Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

1985

Models to predict the performance of axial flow carbon dioxide absorptive canisters.

Yarborough, Joseph Earl Jr.

http://hdl.handle.net/10945/21446

Downloaded from NPS Archive: Calhoun

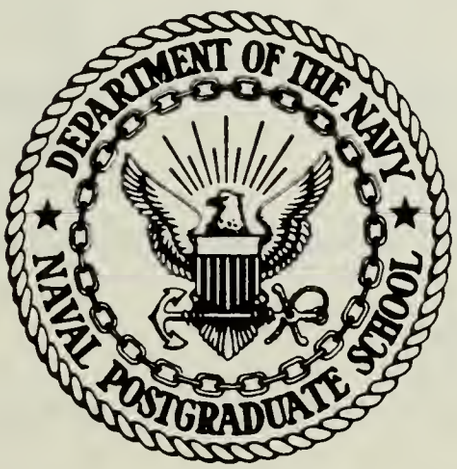
http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943

NAVAL POSTGRADUATE SCHOOL
Monterey, California



THESIS

MODELS TO PREDICT THE PERFORMANCE OF AXIAL FLOW CARBON DIOXIDE ABSORPTIVE CANISTERS

by

Joseph Earl Yarborough, Jr.

March 1985

Thesis Advisor: Charles W. Hutchins Jr.

Approved for public release; distribution unlimited.

T223038

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Models to Predict the Performance of Axial Flow Carbon Dioxide Absorptive Canisters		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis March 1985 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Joseph Earl Yarborough, Jr.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		12. REPORT DATE March 1985 13. NUMBER OF PAGES 80
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Carbon Dioxide, Absorption, Linear Regression, Data Analysis, Cross Validation

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Models were developed using the techniques of multiple regression to predict the time at which the carbon dioxide concentration in a gas stream that exits a canister, exceeds the physiological limit for human respiration. Additional

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

#20 - Abstract - (CONTINUED)

models were developed to predict canister efficiency as a function of various geometric and environmental parameters. Simple cross validation was performed in both cases to provide a measure of model applicability.

Approved for public release; distribution unlimited.

Models to Predict the Performance of Axial Flow Carbon Dioxide Absorptive Canisters

by

Joseph Earl Yarborough, Jr.
Lieutenant Commander, United States Navy
B.S., University of South Carolina, 1973

Submitted in partial fulfillment of the
requirements for the degree of

MASTERS OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL

March 1985

1227
C.1

ABSTRACT

Models were developed using the techniques of multiple regression to predict the time at which the carbon dioxide concentration in a gas stream that exits a canister, exceeds the physiological limit for human respiration. Additional models were developed to predict canister efficiency as a function of various geometric and environmental parameters. Simple cross validation was performed in both cases to provide a measure of model applicability.

TABLE OF CONTENTS

I. INTRODUCTION ---------- 6

A. BACKGROUND ---------- 6

B. PHYSIOLOGICAL EFFECTS OF CARBON DIOXIDE ----- 9

C. CHEMICAL PROCESS OF CO_2 ABSORPTION ---------- 12

D. PARAMETERS OF CANISTER CO_2 ABSORPTION ------- 14

E. EXPERIMENTAL DESIGN ---------- 19

II. MODEL DEVELOPMENT ---------- 21

A. THEORY OF REGRESSION ---------- 24

B. CASE ANALYSIS ---------- 30

C. PROCEDURES FOR SUBMODEL SELECTION ---------- 33

D. BUILDING THE MODEL ---------- 35

E. INTERPRETING THE RESULTS ---------- 60

III. CONCLUSIONS ---------- 66

APPENDIX A: DATA ---------- 68

APPENDIX B: DATA STATISTICS ---------- 76

LIST OF REFERENCES ---------- 78

INITIAL DISTRIBUTION LIST ---------- 80

I. INTRODUCTION

The Navy Experimental Diving Unit (NEDU) and the Naval Coastal Systems Center (NCSC) in Panama City, Florida are actively engaged in improvement of the Navy diving program. Numerous tests have been conducted, both manned and simulated, to assess the design considerations for axial flow canisters used to remove excess carbon dioxide from a diver's breathing medium.

In May, 1983, NCSC published a technical manual entitled "Design Guidelines for Carbon Dioxide Scrubbers," which was the result of a large series of experiments conducted at NCSC to isolate the effects of various parameters on canister performance [Ref. 1]. The data from these experiments was obtained and is provided as Appendix A to this report.

The objectives of this thesis are:

1) to provide background for the development of carbon dioxide scrubbing canisters,

2) to present a review of the work of Nuckols, Purer, and Deason in developing Reference 1,

3) to utilize the techniques of multiple regression analysis to identify those parameters of interest in predicting the useful life of a canister, and

4) to develop a model that will be useful in predicting the performance of axial flow carbon dioxide absorptive canisters.

A. BACKGROUND

As early as 1878, when H.A. Fleuss, a British diver, utilized a workable solution of caustic potash to remove carbon

dioxide in a closed circuit self-contained underwater breathing apparatus (SCUBA), chemical absorbing agents were being investigated for removal of carbon dioxide (CO_2) from a breathing medium [Ref. 1: pp. 1-9]. Although the worldwide development of mixed-gas underwater breathing apparatus (UBA) dates from 1912, its development in the U.S. Navy followed World War II. Mixed-gas UBA represents systems that employ a lightweight mixed-gas supply and a diver-worn gas recirculation system to remove carbon dioxide.

With open circuit SCUBA, diving depth and duration are sharply restricted by the low efficiency of gas utilization resulting from the complete discharge of exhaled gases. Approximately five percent of oxygen consumed is actually utilized by the diver at the surface and this percentage decreases with increasing depth [Ref. 2: pp. 11-1]. Consequently, in order to conserve gas supply and extend underwater duration, it is essential to improve this efficiency of gas utilization. This is accomplished by recirculating the diver's breathing medium for reuse, removing the CO_2 produced by metabolic action in the body.

In 1965, Lcdr M.W. Goodman, MC, USN, writing in research report 3-64 for the Navy Experimental Diving Unit, addressed the quantitative considerations of design and performance of cylindrical canisters used to scrub CO_2 in a closed circuit system [Ref. 3]. He stated that research and developmental efforts related to the problem of CO_2 removal form the gaseous atmosphere of mobile closed-circuit life-support systems have

been devoted almost exclusively to submersibles and aerospace vehicle crew modules. There had been no fruitful application of procedure or hardware to the diving community, which needed compact and streamlined packages to interface with a non-propelled diver. Goodman's work was essentially an extension of the work of H.W. Huseby [Ref. 4] and G.J. Duffner [Ref. 5] who developed design criteria for a SCUBA carbon dioxide removal canister presuming satisfactory performance for 180 minutes at 30 feet and 30 minutes at 180 feet. The problem that Goodman addressed was to determine the factors which govern the function of SCUBA CO_2 absorbent canisters, and the quantitative manner of their interdependence. Utilizing data obtained from dives using swimmers and experiments using a mechanical respirator, he was able to relate breathing resistance, duration of useful canister life and efficiency of absorbent utilization, all to canister geometry.

This development of design considerations for the construction of CO_2 absorbent canisters has been an ongoing process at both the Navy Experimental Diving Unit and the Naval Coastal Systems Center, Panama City, Florida. The most recent work has been conducted by M.L. Nuckols, A. Purer, and G.A. Deason of Naval Coastal Systems Center.

Starting in late 1981, they started a series of controlled laboratory tests to isolate the effects of environmental and geometric parameters on canister absorption efficiency. Their ultimate goal was to develop a manual which would enable

engineers to predict the performance of prototype designs of axial flow canisters. This goal was realized with the publication of NCSC Technical Manual 4110-1-83 [Ref. 1]. This manual presents design data and guidelines to predict the performance of axial flow carbon dioxide canister designs using alkali metal hydroxide absorbers. In addition, methods are developed to evaluate efficiencies of different carbon dioxide absorbents in order to characterize the absorption capability of a soda-lime type absorbent presently used by the U.S. Navy.

B. PHYSIOLOGICAL EFFECTS OF CARBON DIOXIDE

This section presents the physiological effects on the body of increased concentrations of carbon dioxide in inhaled gases and is derived from the NCSC TECHMAN [Ref. 1: pp. 2-7]. The process of breathing is characterized by the inhalation of oxygen and the exhalation of a nearly equal volume of CO_2. The ratio of CO_2 produced to oxygen consumed is called the respiratory quotient (RQ) and is assumed to be 0.85 for purposes of canister design. Various respiratory quotients can be derived from Table 1 [Ref. 1: p. 3] for various swim speeds. Respiratory minute volume (RMV), is defined as the volume of gas moved in and out of the lungs per minute, while values for CO_2 produced are based on the density of CO_2 at standard temperature and pressure.

In a closed diving system where CO_2 is included in the inhaled gas, the partial pressure of CO_2 in the blood increases

TABLE 1

O_2 CONSUMPTION AND CO_2 PRODUCED
FOR VARIOUS SWIM SPEEDS

Swim Speed (kt)		O_2 Consumption (lpm)	RMV (lpm)	CO_2 Produced g/min	lpm	% CO_2
Rest	Low	0.26	7.0	0.44	0.221	3.57
	Mean	0.34	9.0	0.57	0.289	3.64
	High	0.42	11.0	0.70	0.357	3.68
0.5	Low	0.67	17.0	1.14	0.576	3.84
	Mean	0.82	20.0	1.41	0.714	4.02
	High	1.00	24.5	1.72	0.870	4.05
0.7	Low	0.96	24.5	1.65	0.835	3.87
	Mean	1.14	25.0	1.96	0.991	4.10
	High	1.35	34.0	2.43	1.228	4.50
0.9	Low	1.13	25.0	1.94	0.982	4.19
	Mean	1.53	37.0	2.75	1.391	4.26
	High	1.90	49.0	3.49	1.768	4.45
1.0	Low	1.34	34.0	2.43	1.219	4.03
	Mean	1.83	44.0	3.36	1.701	4.38
	High	2.26	55.0	4.33	2.190	4.51
1.2	Low	1.87	45.0	3.43	1.738	4.39
	Mean	2.50	60.0	4.79	2.425	4.60
	High	3.03	66.0	5.99	3.030	5.21

causing the body to increase the rate of respiration. Excessive amounts of CO_2 in an inhaled gas, based on partial pressure, can result in toxic effects dependent on exposure times. Figure 1 depicts the physiological effects of CO_2 for different concentrations and exposure periods [Ref. 1: p. 5]. The fractional CO_2 concentration in a gas stream (F_{CO_2}) can be defined as

$$F_{CO_2} = [(\dot{V}_{O_2} \times RQ) \div (Q \times P_B)] \div 28.32$$

where $\dot{V}_{O_2}$ is the diver's oxygen consumption rate in liters per minute, RQ is equal to 0.85, Q is the volumetric flow

rate to the canister in cubic feet per minute, and P_B is the environmental pressure in atmospheres absolute.

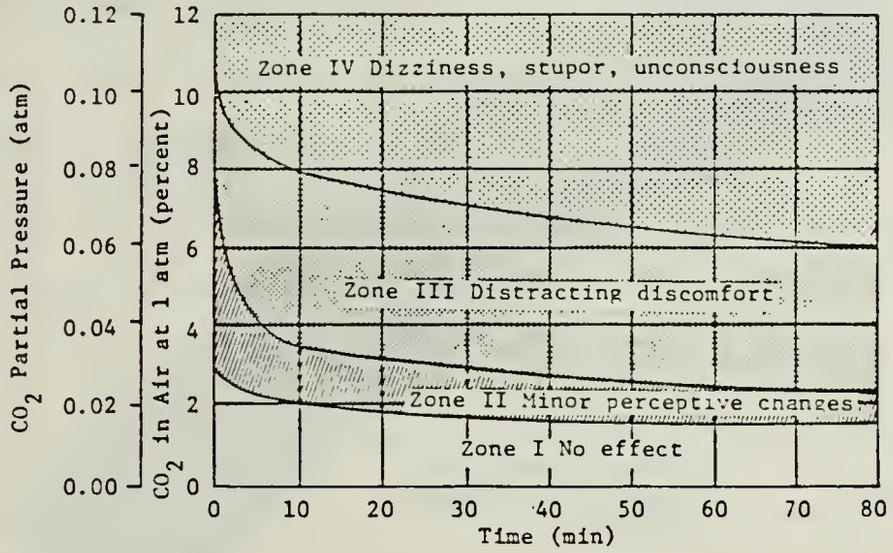


Figure 1. Physiological Effects of CO

Since the units for expressing partial pressure are usually mmHg, the partial pressure of CO_2 (P_{CO_2}) can be computed by

$$P_{CO_2} = F_{CO_2} \times P_B \times 760 \ .$$

Figure 2 [Ref. 1: p. 6] shows CO_2 tolerance zones as a function of percentage of CO_2 and depth at partial pressures taken from Figure 1 for a 1-hour exposure period. The physiological effects of CO_2 depend upon its partial pressure in the breathing gas and thus in the blood. As diving depth increases, the percentage of CO_2 in the breathing gas that can be tolerated decreases.

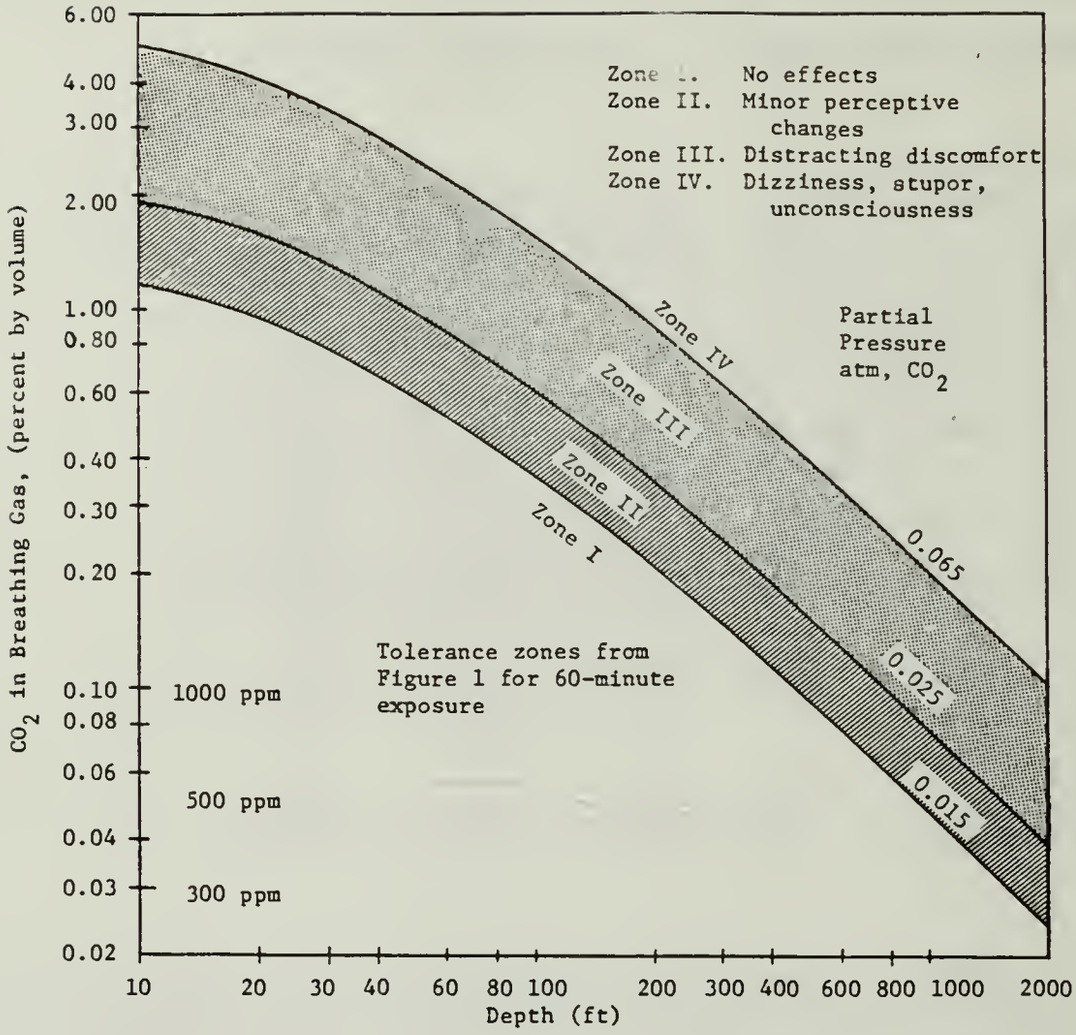


Figure 2. CO Tolerance Zones

In closed-circuit dive systems, all exhaled gases can be rebreathed by passing the exhaled gas through a canister containing a chemical CO_2 absorber before the gas is inhaled. Oxygen that is used in the process of metabolism is replaced by oxygen from a bottle carried by the diver.

C. CHEMICAL PROCESS OF CO_2 ABSORPTION

The CO_2 absorbent used throughout the studies conducted at Naval Coastal Systems Center was High Performance (HP)

Sodasorb, a registered trademark name [Ref. 6]. It consists of calcium hydroxide, 14 to 19 percent moisture by weight, less than five percent sodium hydroxide, potassium hydroxide or barium hydroxide, and ethyl violet as a sensitive acid/base indicator. Ethyl violet is added to the composition to indicate whether the absorbent chemical has been used or not. As the CO_2 is absorbed, the resulting pH change causes the granules to turn from colorless to blue.

The actual process of CO_2 removal by Sodasorb is a chemical reaction. The CO_2 forms carbonic acid which combines with hydroxide to form sodium carbonate and water. Sodium carbonate then reacts with lime to form calcium carbonate. The entire process is characterized by the following equations [Ref. 7: p. 58]:

$$CO_2 + H_2O \rightleftarrows H_2CO_3$$

$$2H_2CO_3 + 2NaOH + 2KOH \rightarrow Na_2CO_3 + 4H_2O + K_2CO_3$$

$$2Ca(OH)_2 + Na_2CO_3 + K_2CO_3 \rightarrow 2CaCO_3 + 2NaOH + 2KOH \ .$$

It is important to note that the chemical reactions will reach their completion provided that the CO_2 laden gas remains inside the canister long enough and adequate mass exchange can be made between the absorbent and the gas. Although water is a byproduct of the above process, water is required to initiate the above reactions. Insufficient moisture will then inhibit the process

while excessive moisture will tend to flood the canister, blocking the critical contact between the CO_2 laden gas and the chemical absorbent particles and also, inhibit or slow down the reaction. Nuckols, Purer, and Deason report that, "the moisture content in the canister can change during its use due to an imbalance in the water content of the incoming and outgoing gas streams. Gas saturated with water normally would result in an increase in moisture; however, if the heat of reaction is not dissipated, high temperature areas will result in the evaporation of water from certain sections of the canister thus rendering such zones ineffective for carbon dioxide removal" [Ref. 7: p. 59].

Although various hydroxides are available for use as chemical absorbents, the U.S. Navy has decided to use absorbents based on calcium hydroxide. Table 2, comprised of data from Reference 1, lists various hydroxides along with various properties. It is of interest to note that while calcium carbonate does not perform as well as lithium hydroxide, its causticity relative to potassium hydroxide is lower.

D. PARAMETERS OF CANISTER CO_2 ABSORPTION

This section is a brief overview of the work of Nuckols, Purer, and Deason [Ref. 1] in identifying the most important or influential variables in the prediction of canister breakthrough time, t. Canister breakthrough time is defined as the time at which the CO_2 concentration in the gas exiting the

TABLE 2

PROPERTIES OF HYDROXIDES

Compound	Chemical Formula	Molecular Weight	CO_2 Capacity $gmCO_2$/100 gm	Solubility in Water gm/100 cc	Causticity
Potassium Hydroxide	KOH	56.1	39.2	115	1.0
Sodium Hydroxide	NaOH	40.0	55.0	109	0.95
Lithium Hydroxide	LiOH	23.9	91.9	13	0.11
Berium Hydroxide	$Ba(OH)_2$	171.4	25.7	5.6	0.05
Calcium Hydroxide	$Ca(OH)_2$	74.1	59.4	0.183	0.0015
Magnesium Hydroxide	$Mg(OH)_2$	58.3	75.5	0.0009	7.0E-7

canister exceeds physiological limits for human respiration. The value of 0.5 percent CO_2 by volume is generally accepted for experimental use. Thus, t provides a measure of a canister's life or effective duration.

The absorption process within a canister involves three separate processes which must work together in order for the canister to perform the scrubbing process. These processes are defined as "mass transfer of CO_2 from the gas stream to the absorbent surface, absorption at the surface, and, finally, chemical absorption" [Ref. 1: p. 12]. In an axial flow canister, if the speed of the gas through the canister is too great, or if the length of the canister is too short, the molecules of CO_2 will not have time to move radially through the

chemical absorbent. This radial movement of gas is a function of the CO_2 concentration, the mass diffusivity of CO_2, the distance from the centerline of the canister to the absorbent surface expressed in terms of canister diameter and absorbent particle size, and the turbulence in the gas stream as a function of gas density and viscosity [Ref. 1].

As discussed earlier, water is necessary for the completion of the absorption process and may be provided in the form of water vapor in the exhaled gas of the diver or in the chemical absorbent itself. Once the chemical absorption process has been started, the reaction rate must be fast enough to keep up with the rate at which CO_2 is arriving. The reaction rate of the process is highly influenced by temperature [Ref. 1].

Table 3 lists the various parameters of interest, which, although not all inclusive, were presented in Reference 1 as having the most influence on canister duration. As the table depicts, canister duration is then a function of fourteen variables:

$$t_B = f(V,T,\rho,\mu,C,H,HA,W,e,L,D,A,D_v,R)$$

This functional relationship was simplified by observing that the absorbent's reaction rate (R), capacity to absorb CO_2 (A), and moisture content (HA) were beyond the control of the user, reducing the number of independent variables to eleven. Next, a dimensional analysis of the resulting relationship was

TABLE 3

PARAMETERS OF CANISTER CO_2 ABSORPTION

Parameter	Identification	Dimensions
t_B	Breakthrough time	t
V	Gas stream velocity	L/t
T	Gas temperature	T
ρ	Gas density	m/L^3
μ	Gas viscosity	m/Lt
C	CO_2 concentration	L^3/L^3
H	Gas stream water vapor content	L^3/L^3
HA	Absorbent water content	L^3/L^3
W	Absorbent mass	m
e	Particle size	L
L	Canister length	L
D	Canister diameter	L
D_v	Mass diffusivity	L^2/t
A	Absorption capacity	m/m
R	Reaction rate	m/t

Basic Units

t = Time m = Mass

L = Length T = Temperature

performed, the theory of which is described in Reference 8. Dimensional analysis is useful when the variables involved in a process are known, but the relationships among the variables are not known. If the variables can be reexpressed by only a few dimensionless groups, then relationships between the groups can be derived which are applicable to all cases within the

range of the dimensionless groups. Since the number of dimensionless groups is less than the original number of variables, considerable time and expense is saved during the process of experimentation [Ref. 8].

Canister efficiency, η, was defined as a ratio of actual breakthrough time, t_B, to theoretical breakthrough time, t_{TH}: $\eta = t_B/t_{TH}$. Theoretical breakthrough time is computed by dividing the absorption capacity of the chemical absorbent by the rate of CO_2 delivered to the canister as follows:

$$t_{TH} = \frac{A\ W}{Q\ C\ \rho_{CO_2}}$$

where:

A = mass of CO_2 absorbed per mass of absorbent;

W = mass of absorbent;

Q = absolute volumetric flow rate;

C = CO_2 volume fraction in gas stream at depth;

ρ_{CO_2} = density of CO_2 at depth for given temperature.

The dimensional analysis resulted in an expression for canister efficiency, η, of:

$$\eta = f(Re_D, T, H, e/D, L/D)$$

where Re_D is the Reynold's number and is defined as:

$$Re_D = \frac{\rho\ \bar{V} e}{\mu} .$$

In other words, efficiency, η, is defined to be a function of five dimensionless groups represented by Reynold's number, temperature, relative humidity, a ratio of absorbent particle diameter to canister diameter, and a ratio of canister length to canister diameter.

E. EXPERIMENTAL DESIGN

The experimental set-up depicted in Figure 3 was utilized by Nuckols, Purer, and Deason to derive the data set relating canister breakthrough time to various parameters [Ref. 9: pp. 3-4]. Helium was utilized as the main ingredient of the inlet gas because of its high thermal conductivity which allowed for quick dissipation of the heat of reaction resulting in a more uniform distribution of temperature across the test canister. Although the diameter of the test cell was varied, it was kept small to insure a constant temperature throughout the absorbent. Relative humidity was varied by passing inlet gas through a bubble tower. Temperature of the test cell was varied from 35 F to 70 F, while weight of the absorbent depended on cell dimensions and packing density. The percentage of CO_2 in the inlet gas was varied over a realistic range for divers, while the flow rate ranged as high as 31,185.35 cubic centimeters per minute measured at the surface. Absolute pressure was varied from one atmosphere (surface) to thirty-two atmospheres (1023 feet dive depth). Various moisture contents of Sodasorb were constructed by adding water or baking the absorbent [Ref. 1].

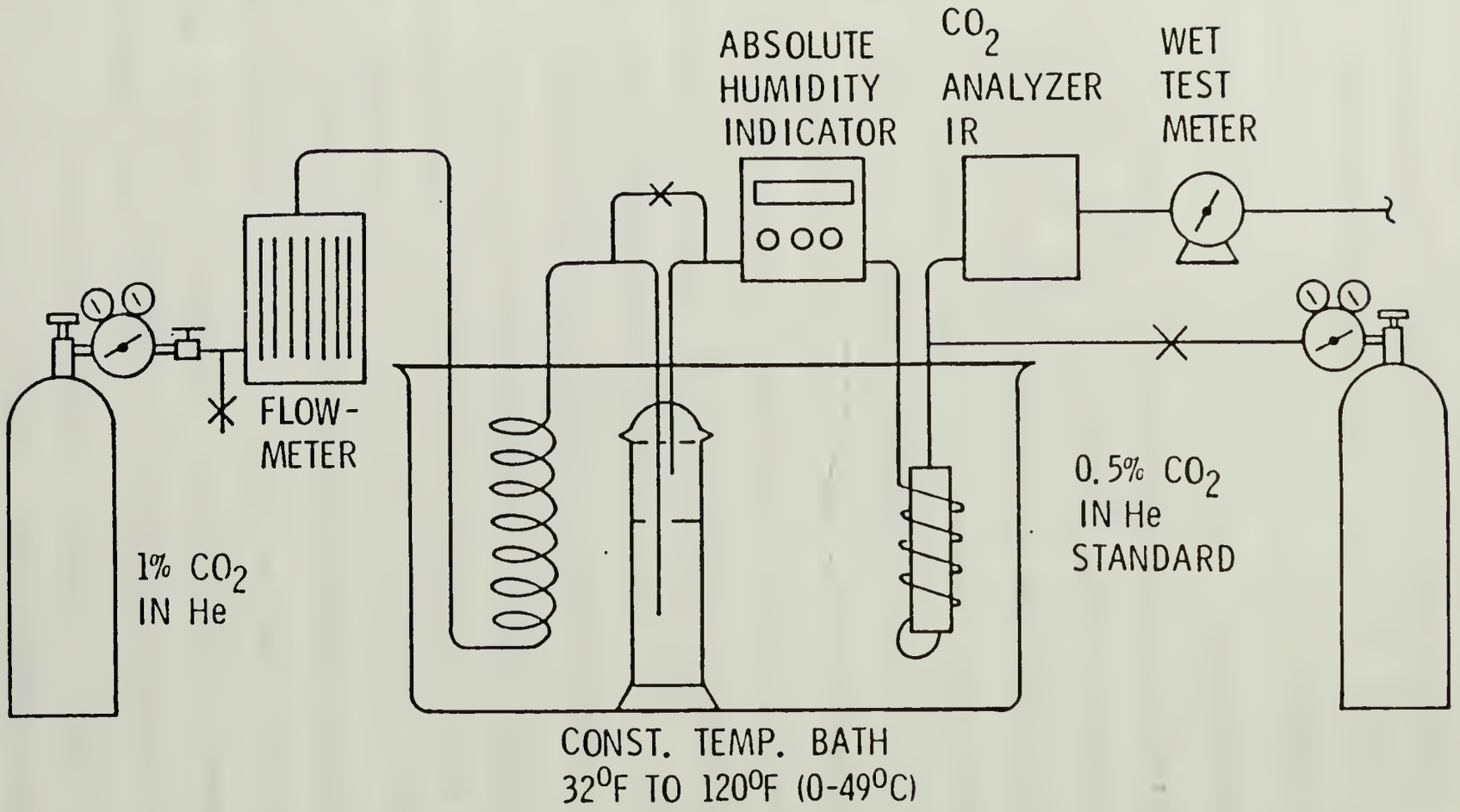


Figure 3. Experimental Design

II. MODEL DEVELOPMENT

The data generated by the experiments of Nuckols, Purer, and Deason are based on steady flow through axial canisters. A diver's respiratory cycle actually represents an intermittent or pulsatile flow, the effects of which are discussed in Reference 10.

The data is in matrix form and is provided in Appendix A. There are 377 rows, representing independent trials of the experiment and 10 columns, representing the 9 independent and 1 dependent variables. The columns from left to right represent relative humidity in percent (RH), temperature in degrees Fahrenheit (TEMP), weight of the chemical absorbent in grams (WT), time to breakthrough in minutes (TIME), canister length in inches (LENTH), canister diameter in centimeters (DIAM), percent by volume of CO_2 in the inlet gas at depth (CO_2), surface rate of flow of the inlet gas in cubic centimeters per minute (FLOW), absolute operating pressure expressed in atmospheres (PRESS), and the hydration level of the chemical absorbent expressed as a percentage (HYDRA).

The ranges of the independent variables are provided in Table 4, while statistics on each variable or column of the matrix, are provided in Appendix B. The data set of Appendix A allows for actual canister breakthrough time, t_B, to be expressed as a function of as many as nine variables, represented as:

$$t_B = f(RH, TEMP, WT, LENTH, DIAM, CO_2, FlOW, PRES, HYDRA) .$$

These nine variables are the same that are utilized in the process of calculating canister efficiency and theoretical canister life to obtain actual canister life as described in the NCSC Techman with the exception of ρ (gas density), μ (gas viscosity), e (absorbent particle diameter), D_v (gas mass diffusivity), and chemical hydration level (HA). Absorbent moisture content or hydration level was assumed to be fixed from the manufacturer, however, samples of varying moisture content were constructed and tested at NCSC and consequently hydration level is included as the independent variable HYDRA.

TABLE 4

RANGES OF THE INDEPENDENT VARIABLES

	MEAN	VAR	MIN	MAX
RH (%)	56.764	2301.2	0	100
TEMP (°F)	66.472	91.8	35	80
WT (gm)	19.994	4049.8	2.05	383.56
LENTH (in)	2.679	1.2	1.078	16
DIAM (cm)	1.489	2.8	0.975	9.446
CO_2 (%)	1.149	1.3	0.031	4.35
FLOW (cc/min)	1039.887	10,965,549.7	23.837	31,185.74
PRESS (ata)	4.202	54.6	1	32
HYDRA (%)	13.386	9.5	0.6	14.1

Although D_v is affected by pressure, it remains fairly constant over the range of depths represented by the data and is therefore excluded [Ref. 1: p. 15]. The absorbent particle diameter remains constant at 0.14 inches and also is excluded. Density and viscosity are dependent on the gas mix, operating pressure and temperature. These variables are required to calculate Reynold's number, which is used to define efficiency, however, it is the attempt of this thesis to dervie a relationship for canister breakthrough time as a function of the data gathered by experimentation.

A multiple regression analysis was performed on this matrix of data with TIME as the dependent or predicted variable. The purpose of this analysis was to examine the relationships existing among the variables and to develop a model that would be useful in making predictions for the dependent variable within the ranges of the independent variables. The most common reasons for performing a regression analysis are to provide a description of the relationships between the predictor variables and the dependent variable and provide a model for predictions of future values. It is important to understand the difference between interpolation and extrapolation when predicting values. Interpolation occurs when the predictor variables are within the multidimensional space observed in construction of the model, while extrapolation represents predictions using values of variables which are outside this space. Regression analysis provides a basis for interpolation, while extrapolations

must be made based on the experience and knowledge of the analyst.

Regression analysis can be separated into two stages, the first of which is aggregate analysis where the intent is to combine the data and develop a model. The second stage is case analysis where the data are used to examine the suitability and correctness of the fitted model. This chapter will present a brief review of regression theory and then develop models to predict effective cannister life, represented by the variable, TIME.

A. THEORY OF REGRESSION

Although many estimation procedures exist for obtaining estimates of parameters in a model, this thesis will use the most common one of least squares. The theory for this development is extracted from References 11 and 12. In simple linear regression, a model is constructed to fit a straight like through a set of data and is of the form: $Y = \alpha + \beta X + \varepsilon$, where Y is the response variable, X the predictor variable, β the slope of the line, α the Y-intercept and ε the error term. The parameters of the line are α and β. Differences in the values predicted by the model and the actual observed values are called statistical errors, or residuals, and have both fixed and random components. The fixed components result from fitting a straight line model to a set of data where the true relationship is curvilinear. Random effects result from measurement

errors, natural variability of the predictors and the effect of not including variables which affect the response. The residuals, ε, are assumed to be normally and independently distributed with zero mean and common variance.

The method of least squares minimizes the residual sum of squares defined by

$$RSS = \sum_{i=1}^{N} \hat{e}_i^2 = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 ,$$

where the hat notation signifies an estimated or predicted value, i.e., $\hat{Y}$ represents the predicted values of the response variable. The sample variance of Y given X can be defined as

$$S^2_{Y|X} = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2/(N-2) ,$$

and represents an unbiased estimate of the population variance. The denominator, N-2, represents degrees of freedom defined as the number of cases minus the number of parameters in the model. The square root of the sample variance of Y given X is called the standard error of regression and will be used later to compare models generated in multiple regression where its definition is similar.

By defining

$$SXX = \sum_{i=1}^{N} (X_i - \bar{X})^2, \quad SYY = \sum_{i=1}^{N} (Y_i - \bar{Y})^2, \quad \text{and}$$

$$SXY = \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}) ,$$

where the bar denotes a sample average, the sum of squares due to regression can be defined as

$$SSR = (SYY - RSS) = (SXY)^2/SXX .$$

In simple regression, the degrees of freedom (d.f.) associated with SSR is one.

The coefficient of determination for multiple regression is defined as $R^2 = SSR/SYY$ and represents a scale free one-number summary of the strength of the relationship between the X and Y of the data. It is easily interpreted as a percentage of variability in the criterion that is explained by the model. It also is the same as the square of the sample correlation between X and Y. R^2 has long been used to compare models of the same data and its value in comparisons will be discussed later.

Multiple regression implies that two or more predictor variables are used to model the response variable. The basic model is: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$ where α and β are unknown parameters, e represents statistical error, Y is the response variable and the X's are the independent variables. The "real world" geometric interpretation of this model breaks down as it is the equation of a K-dimensional hyperplane in (K+1) dimensional space.

Analysis of variance (ANOVA) provides a method whereby the fit of various models to the same set of data can be

compared. The F statistic is defined as F = MSR/RMS, where MSR is the regression mean square obtained by dividing SSR by d.f. and RMS is the residual mean square obtained by dividing RSS by d.f. It compares the variation due to regression to variation due to error. The F statistic is used to test the hypothesis that all of the regression weights in the model are equal to zero. If the calculated F value exceeds the tabled value with the appropriate degrees of freedom, then one can be at least $100(1-\alpha)\%$ sure that at least one of the predictors is significant [Ref. 12]. The significance level is represented by α.

The coefficient of determination, R^2, is now defined as $R^2 = ((SYY-RSS)/SYY) = (SSR/SYY)$ and is the same as the square of the multiple correlation coefficient between Y and the X's. Most importantly, R^2 is the square of the maximum correlation between Y and any linear function of the X's, i.e., the predicted values of Y. Table 5 provides a summary of ANOVA for multiple regression where K is the number of predictor variables and N is the number of cases or trials.

TABLE 5

ANOVA FOR MULTIPLE REGRESSION

Source	SS	d.f.	MS	F
Regression $(x_1, X_2, \ldots, X_k)$	SSR	K	MSR = SSR/k	MSR/RMS
Residuals (error)	RSS	N-(K+1)	RMS = RSS/(N-K-1)	
Total	SST = SYY = RSS + SSR	N-1		

Since the calculation of R^2 is dependent on the data set from which the model was constructed, R^2 associated with the development sample will tend to be larger than R^2 for the population. This has led to the definition of adjusted R^2, R_A^2, which is an attempt to correct R^2 to more closely fit the population. Adjusted R^2 is defined as

$$R_A^2 = R^2 - ((P(1-R^2))/(N-P-1)) ,$$

where N is the number of cases and P is the number of predictor variables in the model [Ref. 13: p. 97].

Lane has defined the terms overfit and shrinkage to explain the facts that R^2 for the equation development sample will be larger than R^2 for the population (overfit), while R^2 developed from using a model to predict over a new sample will be lower than R^2 for the equation development sample (shrinkage) [Ref. 14]. His study proposes three shrinkage estimation formulas identified by their authors as follows:

Wherry: $$R_A^2 = 1 - (1-R^2)[(N-1)/(N-P-1)]$$

Nicholson-Lord: $$R_A^2 = 1 - (1-R^2)[(N-1)/(N-P-1)][(N+P-1)/N]$$

Darlington: $$R_A^2 = 1 - (1-R^2)[(N-1)/(N-P-1)][(N+1)/N] \times [(N-2)/(N-P-2)]$$

where N is the number of cases, P is the number of predictor variables, R^2 is the coefficient of determination for the

equation development sample and R_A^2 represents R^2 adjusted for the population. The Wherry formula can be shown to be equivalent to the formula proposed earlier in Reference 13 for adjusted R^2. The three formulas are, from top to bottom, increasingly conservative. The Wherry formula was developed to estimate the value of R^2 to be expected if weights from a given sample were applied to a large number of samples from the same population. Nicholson and Lord noted that the Wherry formula actually estimates R^2 for the population; i.e., the result expected if population weights were applied to the entire population rather than the result of applying a sample equation to the population. Both the Nicholson-Lord and Darlington equations are derived from equations developed for residual mean square and estimate R^2 resulting from the application of a sample equation to the population [Ref. 14].

In simple regression, a t-statistic can be computed as $t = \hat{\beta}/S_{y|x}$, and is used to test the null hypothesis that $\hat{\beta}$, or the slpe, is equal to zero. It is observed that $t^2 = F$ when d.f. = 1. In multiple regression, the t-statistic is used to test the null hypothesis that any one regression weight or coefficient estimate is equal to zero. The t-statistic is calculated by dividing the estimate of the coefficient by its standard error. This statistic is then compared to the tabled t value with d.f. equal to N-K-1.

In order to place any validity on the results obtained from ANOVA techniques, various assumptions are made concerning

the data and distribution of the error term. These assumptions are discussed informally by Amick and Walberg [Ref. 15] and more formally by Weisberg [Ref. 11] and Younger [Ref. 12]. There are basically five of these assumptions which will be presented next.

1. The model is correctly specified. Every variable (both dependent and independent) is in its proper form. Linear regression requires that the relationship of the response variable to the set of predictor variables be linear. If the relationship is actually curvilinear, then transformations of the data must be used to transform a theoretically nonlinear model to linear form.

2. The data are representative of the range over which predictions are to be made.

3. Each observation or trial can be considered to be independent of the others.

4. The errors in the model are assumed to be normally and independently distributed with zero mean and common variance.

5. The distribution of the possible values of the response variable must be normal about the fitted line for all values of the predictor variables.

These assumptions, as they apply to the data set predicting canister life, will be discussed later in this chapter. The actual technique of checking for inadequacies of a model is called case analysis and is discussed next.

B. CASE ANALYSIS

The analysis of the correctness and suitability of a model is called case analysis. In order to understand the technique and to place the author and reader on common ground, it is necessary to define various terms used as well as notation.

Weisberg defines studentized residual, r_i, as the residual divided by an estimate of its standard deviation. The advantage of studentized residuals (SRESID) over residuals is that that standard deviation of the $r_i = 1$, for all i. The studentized residual is also used in the calculation of a t-statistic to be used in a test for outliers. The t-statistic is calculated by

$$t_i = r_i((N-K-1)/(N-K-r_i^2)^{1/2}$$

where N is the number of trials and K the number of predictor variables in the model. If the calculated statistic exceeds the table value of critical values for the outlier's test, given significance level α with degrees of freedom N, then the hypothesis that the case is not an outlier is rejected with confidence $100(1-\alpha)\%$. Table D of Reference 11 is a table of critical values for the outlier's test [Ref. 11].

Outliers represent data points that lie beyond the fitted equation and consequently will not be predicted with much accuracy by the model. They are easily identified in case analysis by their large residual or studentized residual and can be confirmed by the outlier's test previously discussed. Outliers can be the result of randomness of the data, experimental error, or they might represent new and unexplainable information. They can not be routinely deleted from the analysis, but must be examined for the source of their aberration.

The studentized residuals defined above represent a scaling of residuals, so that cases which are located far from the centroid of the data space result in increased residuals, while cases closer to the centroid result in decreased residuals. Cook's distance is defined using studentized residuals and is used as a measure of the influence a single point has on the development of the model. The size or magnitude of Cook's distance will be determined by the magnitude of the studentized residual and the distance of the point from the centroid of the data space. A data point that exerts a large influence on a model or fitted equation might have a small residual as it essentially pulls the equation towards it, but can be identified by Cook's distance due to its position outside the centroid. Highly influential cases should be examined in an attempt to identify the source or cause of their uniqueness and, if deleted, will result in a lower value of R^2. However, more confidence can be placed in the model in terms of its future application [Ref. 11].

The single most important plot in case analysis is the plot of residuals with the predicted values of the dependent variable, Y. Weisberg recommends using studentized residuals in the plot because each r_i has zero mean and variance of 1 so that 95% of the r_i should fall within ±2 and 99% should be contained within ±3 [Ref. 11]. The plot should exhibit no characteristic trends, i.e., should be a random scatter of points about the axes. Such a plot would support the assumptions of linearity of the model and common variance for the errors.

The standardized residual (ZRESID) for the i^{th} case is defined by Norusis as the residual divided by the sample standard deviation of the residuals [Ref. 13]. It is useful in examining the assumption of normality of the errors of the model. As with the studentized residuals, 95% should lie between ±2 and 99% between ±3. Additionally, a normal probability plot can be used as a check to see if the residuals represent a homogeneous sample from a normal distribution. The closer the plot resembles a straight line, the more certain one is that the distribution of residuals is normal. If the distribution is not normal, then a possible remedy is a transformation of the dependent and/or response variable [Ref. 11].

C. PROCEDURES FOR SUBMODEL SELECTION

Once the set of predictor variables has been defined and appropriate transformations of the data have been made, the next step is to select that combination of variables that provides the best model. The number of possible models to examine is obviously based upon the number of predictor variables available. If that number is large enough, a factor analysis is appropriate to construct a smaller set of factors that are combinations of the variables and essentially reduce the data set to the relevant dimensions of that set.

With the computer it is possible to search through a large number of variables to select subsets of those variables as possible models. One method of subset selection is called

stepwise regression for which three algorithms exist; forward, backward, and stepwise. The main advantages of these algorithms are that they are fast and relatively easy to compute. They do not provide a "best" model, but examine subsets following a systematic technique. The three methods of stepwise regression are discussed by Weisberg [Ref. 11] and proceed as follows.

The forward selection algorithm first selects the independent variable with the highest sample correlation with the response variable. The next independent variable added is chosen by the following three equivalent criteria: 1) among the remaining variables, it has the highest correlation with the dependent variable, adjusting for the independent variable already chosen; 2) its addition will yield the largest increase in R^2 of the remaining variables; and 3) its t-statistic is largest of the remaining variables. Forward selection can be stopped by the following rules: 1) stop when a limit on the number of independent variables is reached; 2) stop when the significance level of the t-statistic for the remaining variables exceeds a pre-set value; 3) stop when the addition of another variable would make the set of predictor variables too collinear. Collinearity is considered by computing a linear regression of the independent variable of interest with the variables already in the model. R^2 for this regression is computed and tolerance is defined as $1-R^2$. If the tolerance is less than some pre-set value, then the procedure stops. The drawback to this method is that a variable selected early

in the model development may become non-significant and yet remain in the final model.

The backward elimination algorithm starts with a model constructed with all of the independent variables and then removes the variables one at a time. The variable removed is the one having the smallest t-statistic. Stopping rules are: 1) stop when the number of independent variables specified are remaining or 2) stop if the significance level of the remaining variables exceeds a pre-set value.

The stepwise procedure is actually a combination of the forward and backward methods. It selects the first variable as the one with the highest sample correlation. From then on, it selects and ejects variables based on entry and removal criteria set by the user and based on the significance level of the t-statistic for the variable.

D. BUILDING THE MODEL

The New Regression package of the SPSS (Statistical Package for the Social Sciences) Batch System was used on the IBM 3033 to perform a regression analysis of the data in Appendix A. The benefits provided by this package are: 1) a compute statement that allows the computation of transformations of the variables to provide a better fit, 2) the ability to use the forward, backward, or stepwise algorithms for model selection, 3) an extensive case analysis and graphis capability, which enable a check of assumptions.

Examination of the correlation matrix of Table 6 reveals that diameter, flow, and weight are all highly correlated. The correlation between diameter and weight is expected because as the diameter of the test cell is increased, the amount of absorbent that the cell can hold, increases. The correlation of flow rate to diameter and weight appears to be a result of the experimental manipulation of flow to represent various rates for different dive profiles. Flow rates were kept relatively low due to the small size of the test canister.

TABLE 6

CORRELATION MATRIX

	RH	TEMP	WT	TIME	LENTH
RH	1.000	-0.032	-0.293	0.127	-0.108
TEMP	-0.032	1.000	0.093	0.122	0.050
WT	-0.293	0.093	1.000	0.081	-0.045
TIME	0.127	0.122	0.081	1.000	0.192
LENTH	-0.108	0.050	-0.045	0.192	1.000
DIAM	-0.322	0.113	0.974	0.062	-0.093
CO_2	0.043	-0.025	-0.015	-0.218	-0.140
FLOW	-0.249	-.091	0.854	-0.086	-0.063
PRESS	0.127	0.157	-0.109	0.084	0.016
HYDRA	-0.041	-0.085	0.060	0.111	-0.007

	DIAM	CO_2	FLOW	PRESS	HYDRA
RH	-0.322	0.043	-0.249	0.127	-0.041
TEMP	0.113	-0.025	0.091	0.157	-0.085
WT	0.974	-0.015	0.854	-0.109	0.060
TIME	0.062	-0.218	-0.086	0.084	0.111
LENTH	-0.093	-0.140	-0.063	0.016	-0.007
DIAM	1.000	0.044	0.850	-0.133	0.071
CO_2	0.044	1.000	-0.025	-0.406	0.023
FLOW	0.850	-0.025	1.000	-0.102	0.041
PRESS	-0.133	-0.406	-0.102	1.000	0.101
HYDRA	0.071	0.023	0.041	0.101	1.000

Making the assumption that the nine independent variables are the predictors most affecting the dependent variable, time,

a regression of time with these nine variables produced an equation of the form

$$Y = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_i X_i$$

with R = 0.234. The independent variables selected were CO_2, LENTH, RH, DIAM, FLOW, HYDRA, and TEMP. Table 7 lists the ANOVA table and variable coefficients generated by the regression program. Examination of the histogram and normal probability plot of studentized residuals reveals departure from normality. A histogram of residuals is provided in Figure 4.

TABLE 7

ANALYSIS OF VARIANCE TABLE

SOURCE	SS	DF	MS	F
GRAND MEAN	25761545.093	1		
REGRESSION	8997912.646	7	1285416.092	16.090
RESIDUAL	29478974.041	369	79888.819	
TOTAL	64238431.780	377	170393.718	

THE SIGNIFICANCE OF REGRESSION = 1.0000
(SIGNIFICANCE: AREA UNDER CURVE FROM 0 TO COMPUTED F)
R SQUARE = .234

	COEFFICIENT		B/SIGMA(B)	CONFIDENCE INTERVAL	
TERM	B	SIGMA(B)	T	LOWER	UPPER
B0	-414.684	133.426	-3.108	-677.112	-152.256
B1	-67.361	12.927	-5.211	-92.787	-41.936
B2	57.751	13.465	4.289	31.267	84.234
B3	1.516	.325	4.661	.877	2.156
B4	120.644	17.218	7.007	86.778	154.509
B5	-.055	.008	-6.571	-.072	-.039
B6	11.945	4.762	2.508	2.579	21.311
B7	3.472	1.542	2.252	.439	6.504

THE THEORETICAL VALUE FOR T AT THE 0.05 LEVEL AND 369 DF = 1.967

$$TIME = f(CO_2, LENTH, RH, DIAM, FLOW, HYDRA, TEMP)$$

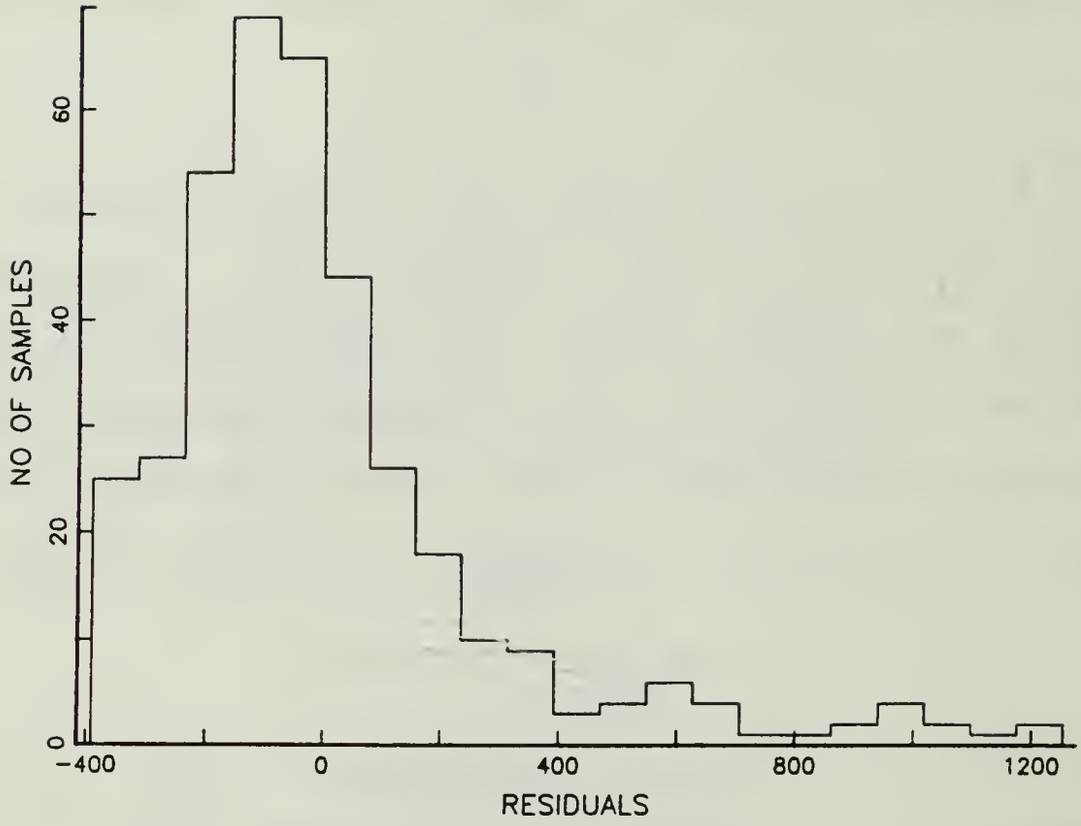


Figure 4. Histogram of Residuals

The plot of residuals versus predicted values in Figure 5, reveals structure and suggests that a transformation of data is required.

Examination of Figure 6 reveals that the distribution of TIME is positively skewed and that the natural log transformation of time, LN TIME, more closely represents a normal distribution. Consequently, the next step was to regress LN TIME with all of the variables. This results in an equation of the form:

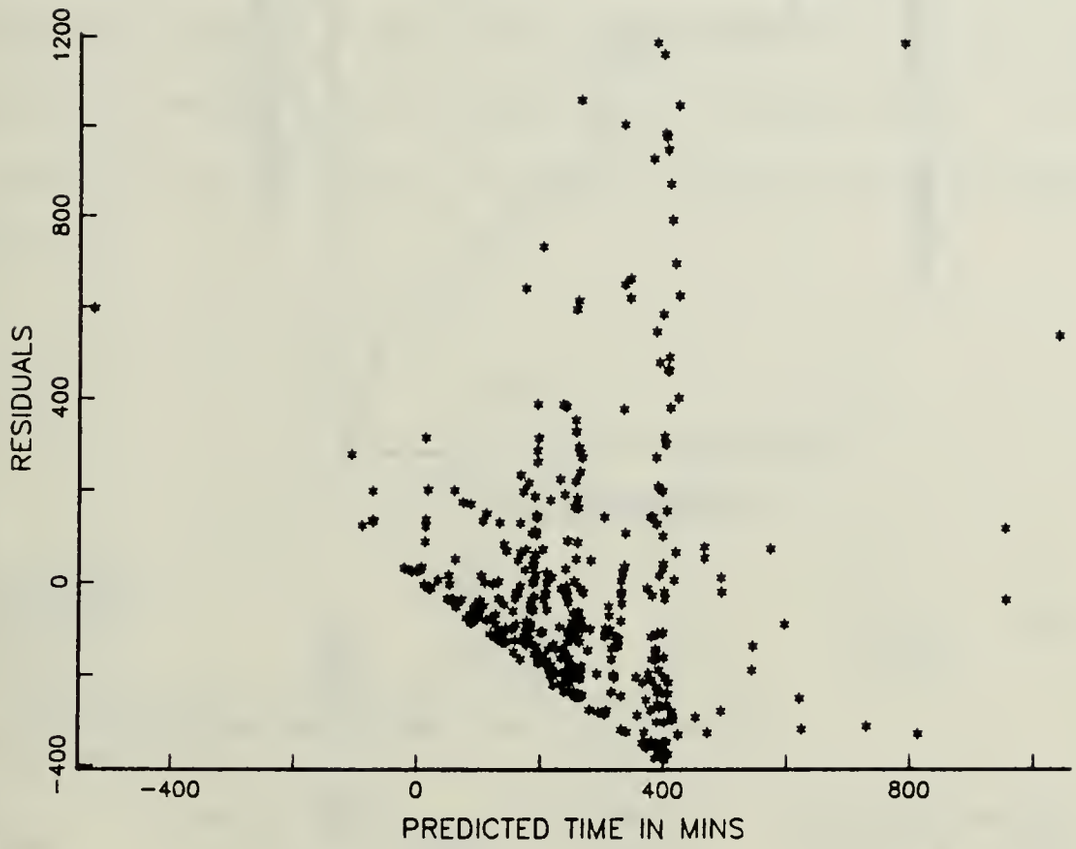


Figure 5. Plot of Residuals vs. Predicted

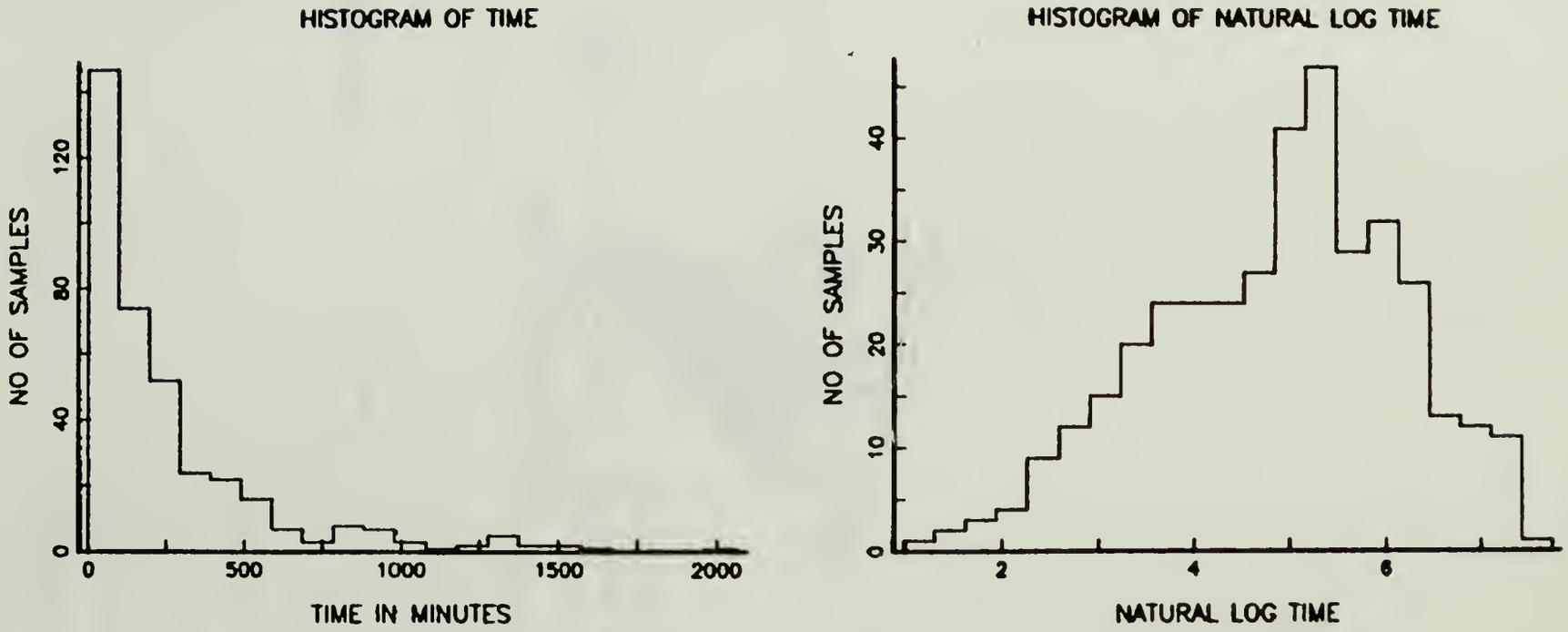


Figure 6. Frequency Distributions

$$Y = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_i X_i ,$$

where the variables selected are HYDRA, LENTH, WT, CO_2, RH, and FLOW. R^2 has reduced to 0.218 as depicted in Table 8 which is the ANOVA table with variable coefficients. Both the histogram of Figure 7 and the normal probability plot have improved to support the assumption of normality of the errors and most of the structure of the residuals has disappeared (Figure 8).

TABLE 8

ANALYSIS OF VARIANCE TABLE

SOURCE	SS	DF	MS	F
GRAND MEAN	8928.089	1		
REGRESSION	137.544	6	22.924	17.203
RESIDUAL	493.058	370	1.333	
TOTAL	9558.691	377	25.355	

THE SIGNIFICANCE OF REGRESSION = 1.0000
(SIGNIFICANCE: AREA UNDER CURVE FROM 0 TO COMPUTED F)
R SQUARE = .218

TERM	COEFFICIENT B	SIGMA(B)	B/SIGMA(B) T	CONFIDENCE INTERVAL LOWER	UPPER
B0	3.613	3.29E-01	10.985	2.966	4.260
B1	.068	1.93E-02	3.528	.030	.106
B2	.161	5.47E-02	2.943	.053	.268
B3	.012	1.82E-03	6.704	.009	.016
B4	-.286	5.24E-02	-5.463	-.389	-.183
B5	.004	1.31E-03	2.836	.001	.006
B6	.000	3.46E-05	-6.034	.000	.000

THE THEORETICAL VALUE FOR T AT THE 0.05 LEVEL AND 370 DF = 1.967

LN TIME = f(HYDRA,LENTH,WT,CO_2,RH,FLOW)

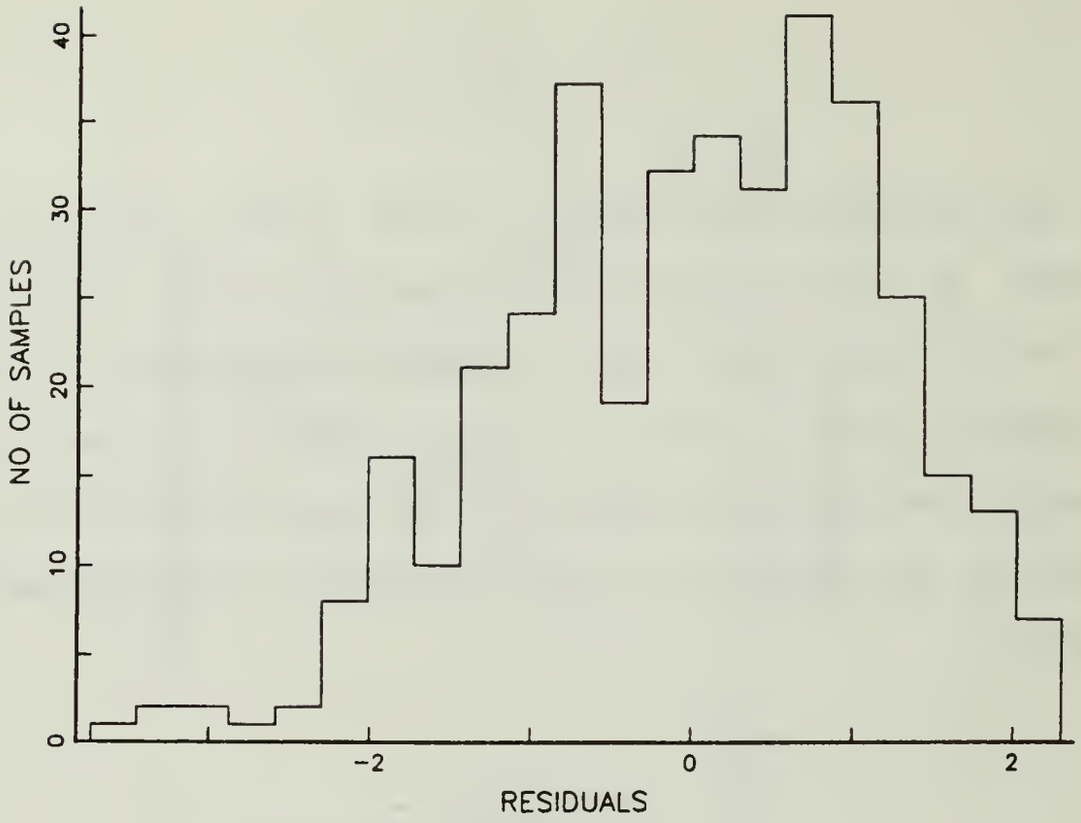


Figure 7. Histogram of Residuals

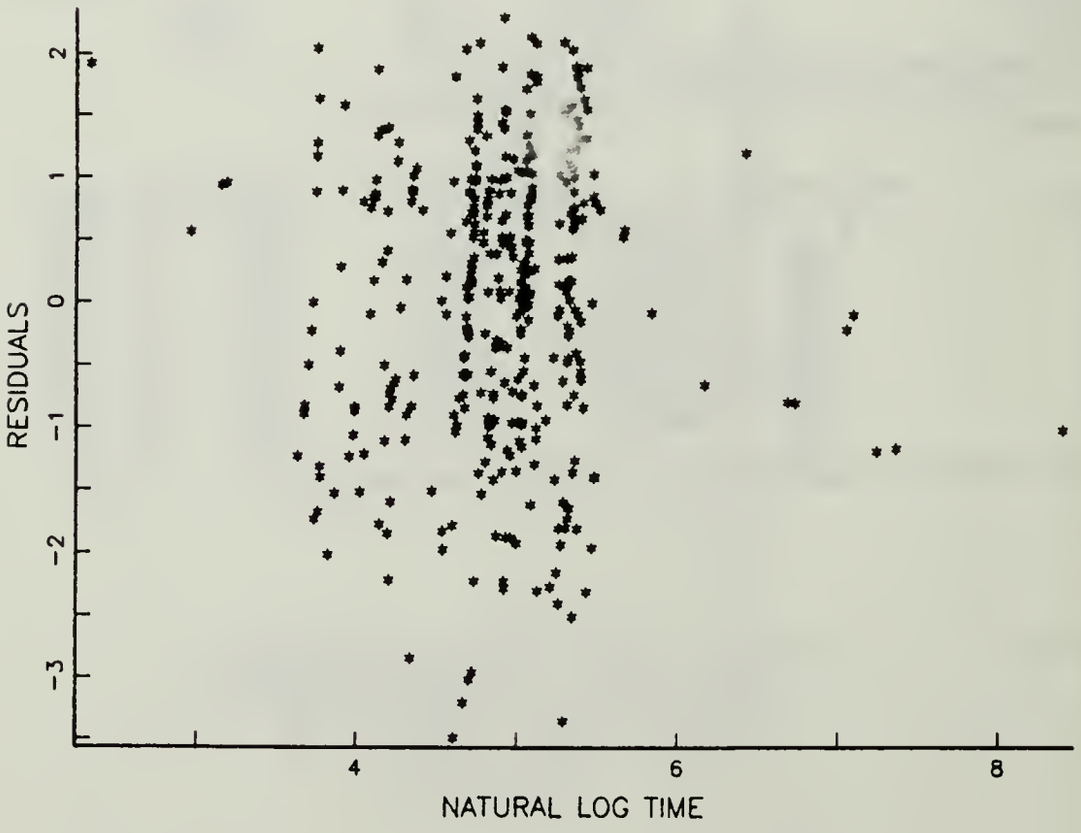


Figure 8. Plot of Residuals vs. Predicted

The current model is then an exponential model of the form

$$Y = \exp(\hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_i X_i)$$

and the relationship between TIME and the predictor variables is curvilinear. The equation is inherently linear as it is possible to use the naturla log transformation to yield a linear form

$$\ell n\ Y = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_i X_i \ .$$

Examination of the case statistics reveals that case number 292 has a large value of Cook's distance of 0.2355 with the next largest value equal to 0.04295. Since its studentized residual is only moderately large, its influence is most likely large and it is deleted for the purposes of further analysis. Case 292 is the case with the highest recorded flow rate and might possibly represent failure of the model to predict over high rates of flow.

Next, a regression of the natural log of time with all of the predictor variables, plus second order terms and natural log terms was run. In order to enable the calculation of a natural log of RH, which at times is equal to zero, 0.0001 was added to all values of RH. Using the stepwise algorithm, the regression yielded an equation of the form

$$\text{LN TIME} = f(\text{LN FLOW}, \text{LN WT}, (CO_2)^2, \text{RH}, \text{TEMP}^2, \text{PRESS}) + \text{constant}$$

with R^2 = 0.880. Table 9 lists the ANOVA table and variable coefficients.

TABLE 9

ANALYSIS OF VARIANCE TABLE

SOURCE	SS	DF	MS	F
GRAND MEAN	8928.089	1		
REGRESSION	554.854	6	92.476	451.708
RESIDUAL	75.748	370	.205	
TOTAL	9558.691	377	25.355	

THE SIGNIFICANCE OF REGRESSION = 1.0000
(SIGNIFICANCE: AREA UNDER CURVE FROM 0 TO COMPUTED F)
R SQUARE = .880

	COEFFICIENT		B/SIGMA(B)	CONFIDENCE INTERVAL	
TERM	B	SIGMA(B)	T	LOWER	UPPER
B0	10.155	1.65E-01	61.371	9.830	10.481
B1	-1.540	3.13E-02	-49.225	-1.602	-1.478
B2	1.583	3.77E-02	42.043	1.509	1.658
B3	-.123	4.65E-03	-26.516	-.132	-.114
B4	.007	5.34E-04	12.616	.006	.008
B5	.000	2.26E-05	7.117	.000	.000
B6	-.016	3.36E-03	-4.627	-.022	-.009

THE THEORETICAL VALUE FOR T AT THE 0.05 LEVEL AND 370 DF = 1.967

$$\text{LN TIME} = f(\text{LN FLOW}, \text{LN WT}, (CO_2)^2, \text{RH}, \text{TEMP}^2, \text{PRESS})$$

All variables are significant at the 0.001 level as is the developed equation. By calculating the t-statistic for an outliers test as defined earlier it is observed that for case 126 with the largest studentized residual of 3.062, the t-statistic is 2.9924. From the table of critical values for the outliers test, t = 3.87 for α = 0.05 and d.f. of 369. This implies that it is not 95% certain that case 126 is an outlier and, therefore, no cases will be deleted. Examination of the histogram of Figure 9 supports the assumption of normality.

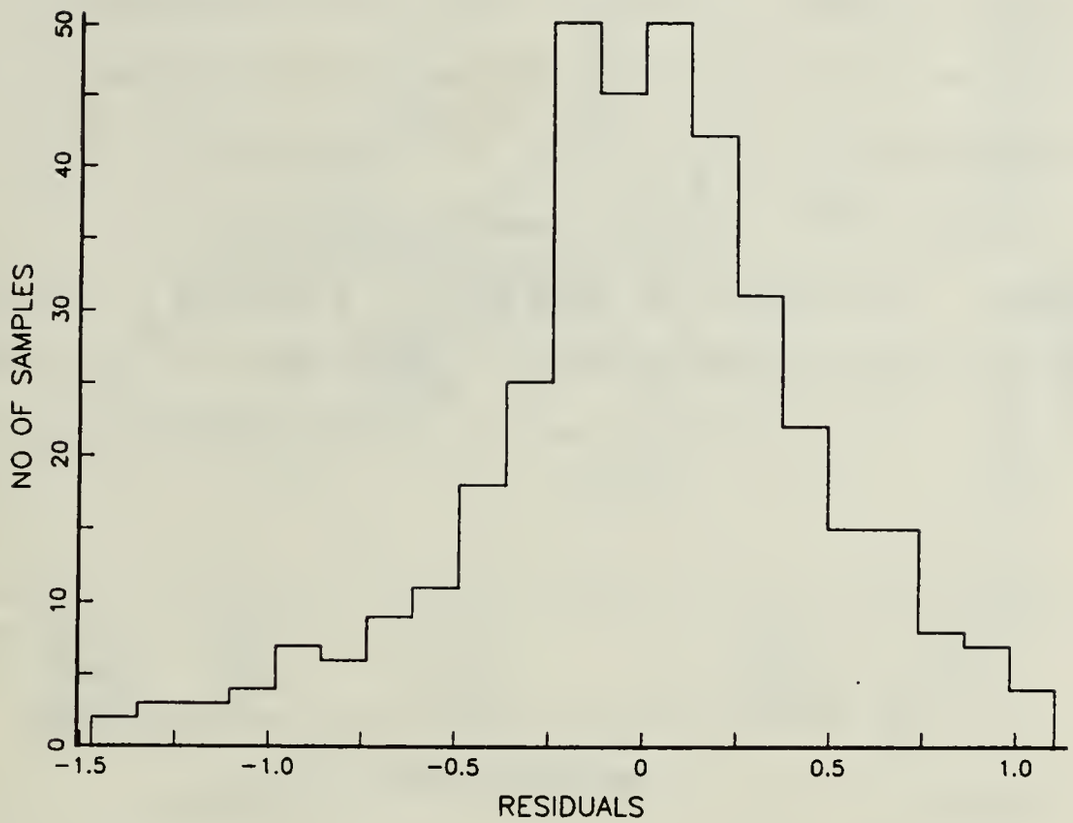


Figure 9. Histogram of Residuals

The residual plot (Figure 10) shows little structure, suggesting the current model is adequate. The model is exponential and can be expressed as

$$Y = \exp(\hat{\beta}_0 + \sum_{i=1}^{N} \hat{\beta}_i X_i) .$$

The set of predictor variables chosen in this final model were selected from a set composed of the original nine independent variables, plus square terms, cross terms, and both inverse and natural log transformations of the independent variables.

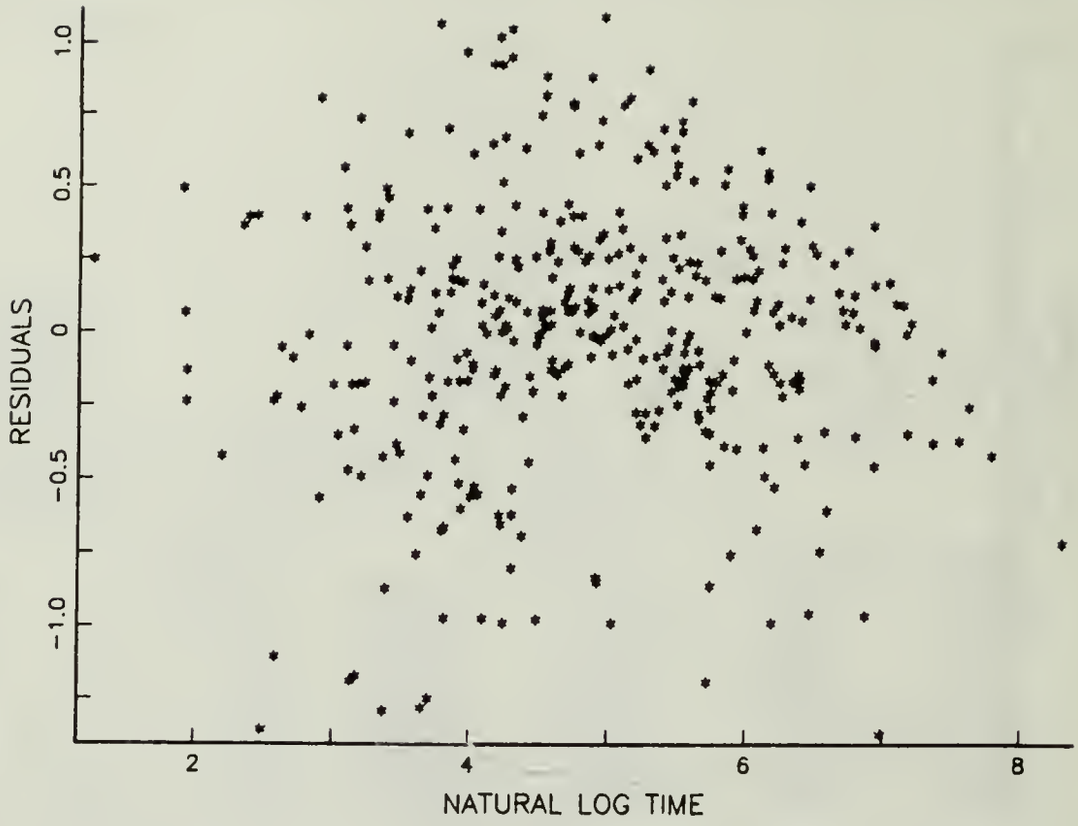


Figure 10. Plot of Residuals vs. Predicted

Using the stepwise procedures of New Regression, it has been possible to adjust the values of tolerance and significance level for the t tests so that the resulting model included only those predictor variables that are significant and not highly correlated. Two questions remain, however, and they are: 1) how large does R^2 need to be before a model represents a meaningful relationship between independent variables and the dependent variable? and 2) how well will a model predict with a new set of data?

The first question can be resolved by returning to the definition of R^2. Based on its definition, the closer R^2 is

to 1.0, the better the fit. However, R^2 is the square of the sample correlation coefficient, R, which measures the strength of the relationship between the predictors and response. The range of R is from minus 1.0 to plus 1.0 with values between minus 1.0 and zero representing inverse relationships and values between zero and plus 1.0, direct relationships. Younger proposes the following scale to represent the strength of the relationship indicated by R [Ref. 12].

-1.0		-.5		0		+.5		+1.0
perfect	moderately strong	moderate	moderately weak		moderately weak	moderate	moderately strong	perfect
		inverse				direct		

For example, from the graph, an R^2 of 0.5625 would translate to an R of 0.75 indicating a moderately strong relationship.

The second question concerns the data dependency of R^2. Adjusted R^2 can be calculated as previously discussed to predict a value of R^2 for the population or the model can be tested by examining how well it predicts with a new set of data. Since a regression model is developed to fit a data set as closely as possible, testing how well a model predicts over this same data set is almost certain to overestimate performance [Ref. 17]. This idea has led to the notion of simple cross-validation, where a data set can be split in half and the model then developed over one half and tested on the other half. The

model can be tested by comparing standard errors or by using the standard errors to calculate R^2 for the equation development sample (R_S^2) and R^2, for the cross-validation sample (R_{CV}^2). The difference of $R_S^2 - R_{CV}^2$ was defined earlier as shrinkage by Lane [Ref. 14].

Based on the above discussion, the data set was split in half using a random number generator and a regression of LN TIME with all variables plus natural log, second order, and inverse terms examined. This resulted in an equation of the form

$$\text{LN TIME} = f(CO_2, \text{LN FLOW}, \text{LN WT}, \text{LN RH}, \text{LN PRESS}, \text{TEMP}) + \text{constant}$$

with $R^2 = 0.91996$. This model can be represented as

$$\text{LN TIME} = \hat{\beta}_0 + \hat{\beta}_1 CO_2 + \hat{\beta}_2 \text{LN FLOW} + \hat{\beta}_3 \text{LN WT} + \hat{\beta}_4 \text{LN RH} + \hat{\beta}_5 \text{LN PRESS} + \hat{\beta}_6 \text{TEMP}$$

or

$$\text{TIME} = \exp(\hat{\beta}_0 + \hat{\beta}_1 CO_2 + \hat{\beta}_2 \text{LN FLOW} + \hat{\beta}_3 \text{LN WT} + \hat{\beta}_4 \text{LN RH} + \hat{\beta}_5 \text{LN PRESS} + \hat{\beta}_6 \text{TEMP})$$

and is exponential in form. Table 10 lists the ANOVA table and variable coefficients generated by the regression program. It is observed from the table that all coefficients are significant at the $\alpha = 0.0001$ level and the standard error of each

coefficient is small. The column labeled BETA, represents the standardized regression coefficients derived by multiplying the variable coefficient by a ratio of the standard deviation of the independent variable to the standard deviation of the dependent variable [Ref. 13]. The actual coefficients do not represent the importance of the variable in the regression equation, since the independent variables are not expressed in the same units. The BETA weights when standardized are dimensionless, however, even then, they do not represent importance since the predictors are intercorrelated.

TABLE 10

ANALYSIS OF VARIANCE TABLE

SOURCE	SS	DF	MS	F
GRAND MEAN	4524.367	1		
REGRESSION	317.780	6	52.963	346.731
RESIDUAL	27.648	181	.153	
TOTAL	4869.794	188	25.903	

THE SIGNIFICANCE OF REGRESSION = 1.0000
(SIGNIFICANCE: AREA UNDER CURVE FROM 0 TO COMPUTED F)
R SQUARE = .920

TERM	COEFFICIENT B	SIGMA(B)	B/SIGMA(B) T	CONFIDENCE INTERVAL LOWER	UPPER
B0	10.317	.259	39.870	9.806	10.828
B1	-.680	.030	-22.340	-.740	-.620
B2	-1.538	.038	-40.552	-1.613	-1.463
B3	1.625	.049	33.064	1.528	1.722
B4	.058	.005	11.687	.048	.068
B5	-.276	.034	-8.053	-.344	-.209
B6	.022	.003	6.750	.016	.029

THE THEORETICAL VALUE FOR T AT THE 0.05 LEVEL AND 181 DF = 1.974

LN TIME = f(CO_2,LN FLOW,LN WT, LN RH,LN PRESS,TEMP)

Returning to the model developed during cross-validation, the normal probability plot and residual plot of Figure 11 support the assumption of normality of errors. The plot of residuals versus predicted values in Figure 12 shows little, if any, structure.

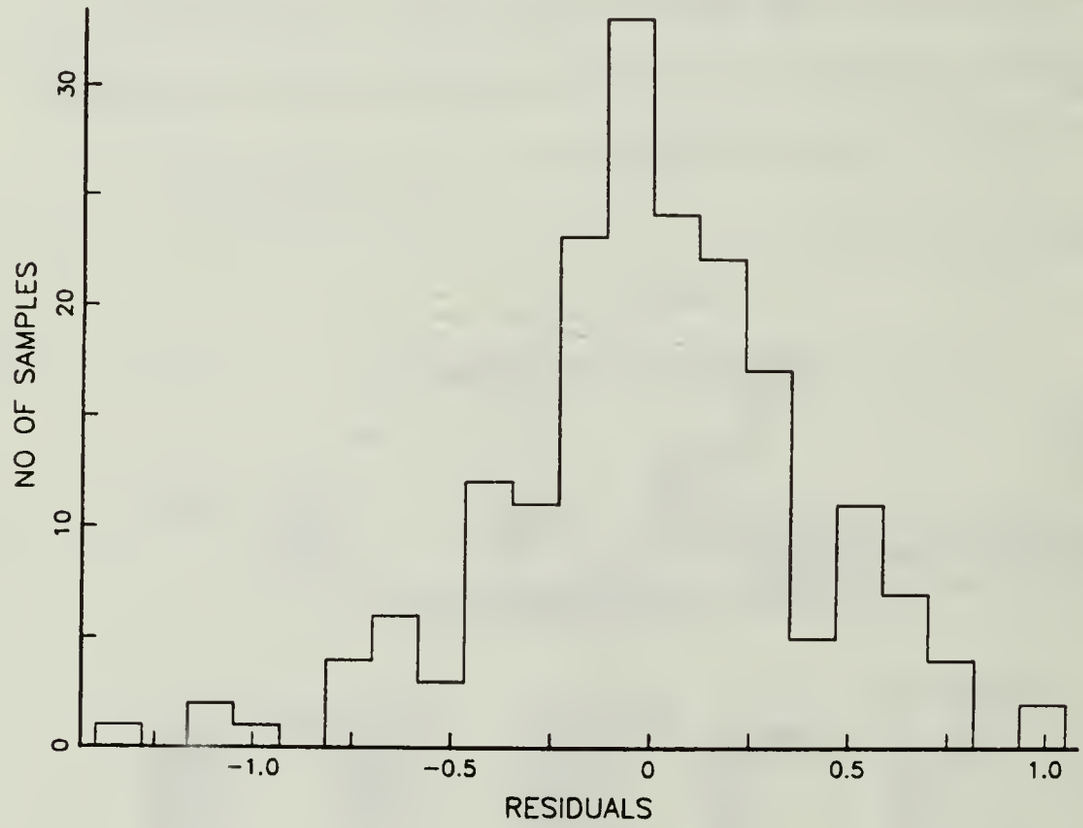


Figure 11. Histogram of Residuals

The plot of actual values versus predicted in Figure 13 supports the assumption of common variance of the residuals. Examination of the ZRESID or standardized residuals reveals that 99% are within ±3 and, therefore, support the contention that residuals or errors are normally distributed.

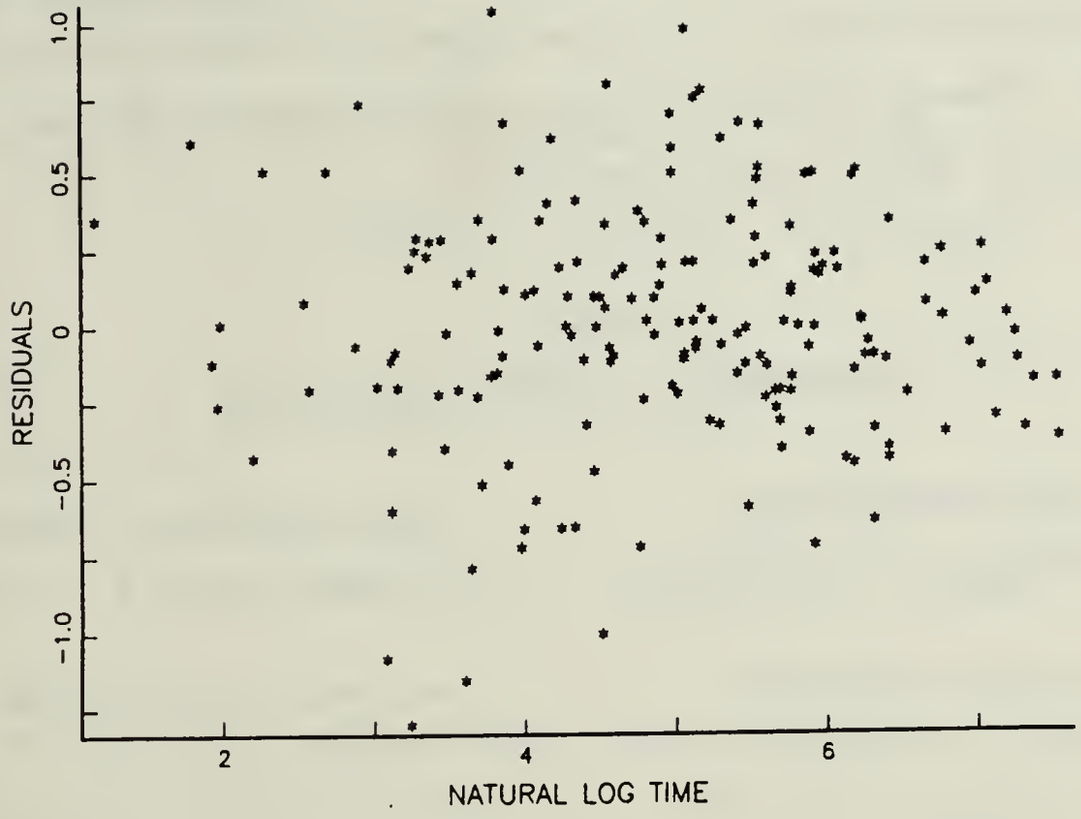


Figure 12. Plot of Residuals vs. Predicted

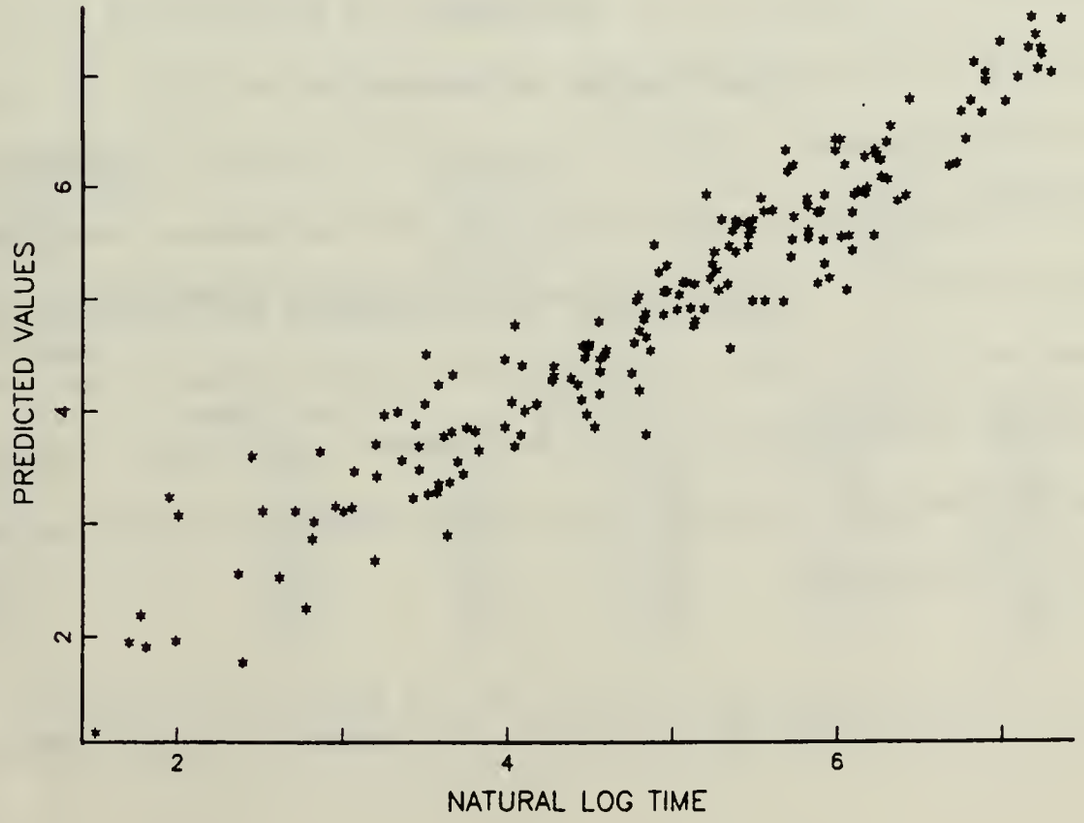


Figure 13. Plot of LN TIME vs. Predicted

The next step was to examine how well this latest model developed with half of the data predicted over the remaining half. Through various calculations as earlier defined, the values in Table 11 were obtained.

TABLE 11

STATISTICS FOR CROSS-VALIDATION

Development Sample	Cross-Validation Sample
Variance of Y = 1.847	Variance of Y = 1.52
Variance of predicted values = 1.699	Variance of predicted values = 1.401
R^2 = 0.9199	R^2_{cv} = 0.8464

These results indicate that although the model does not predict as well over a new set of data as is expected, the obtained R^2 of 0.852 supports the contention of a strong relationship between the predictor variables and the dependent variable.

An additional model was developed using the same half of the data as before, but the model was restricted to having only natural log transformations of the nine independent variables. The New Regression package using the stepwise algorithm selected a model of the form

LN TIME = f(LN FLOW,LN WT,LN CO_2,LN PRESS,LN RH,LN TEMP)

or more appropriately

$$\text{LN TIME} = \hat{\beta}_0 + \hat{\beta}_1 \text{ LN FLOW} + \hat{\beta}_2 \text{ LN WT} + \hat{\beta}_3 \text{ LN } CO_2 + \hat{\beta}_4 \text{ LN PRESS} + \hat{\beta}_5 \text{ LN RH} + \hat{\beta}_6 \text{ LN TEMP})$$

Setting $\hat{\beta}_0 = \ell n\ K$, where $K = \exp(\beta_0)$, the above equation can be transformed to

$$\text{TIME} = K(\text{FLOW})^{\hat{\beta}_1}(\text{WT})^{\hat{\beta}_2}(CO_2)^{\hat{\beta}_3}(\text{PRESS})^{\hat{\beta}_4}(\text{RH})^{\hat{\beta}_5}(\text{TEMP})^{\hat{\beta}_6}$$

which is a multiplicative model. The ANOVA table of Table 12 demonstrates that all coefficients remain significant at the 0.0001 level.

TABLE 12

ANALYSIS OF VARIANCE TABLE

SOURCE	SS	DF	MS	F
GRAND MEAN	4524.367	1		
REGRESSION	319.147	6	53.191	366.335
RESIDUAL	26.281	181	.145	
TOTAL	4869.794	188	25.903	

THE SIGNIFICANCE OF REGRESSION = 1.0000
(SIGNIFICANCE: AREA UNDER CURVE FROM 0 TO COMPUTED F)
R SQUARE = .924

	COEFFICIENT		B/SIGMA(B)	CONFIDENCE INTERVAL	
TERM	B	SIGMA(B)	T	LOWER	UPPER
B0	6.128	.713	8.589	4.720	7.535
B1	-1.551	.037	-41.886	-1.624	-1.478
B2	1.623	.048	33.874	1.528	1.717
B3	-1.609	.069	-23.160	-1.746	-1.472
B4	-1.684	.084	-19.997	-1.850	-1.518
B5	.060	.005	12.290	.050	.069
B6	1.234	.173	7.143	.893	1.575

THE THEORETICAL VALUE FOR T AT THE 0.05 LEVEL AND 181 DF = 1.974

$$\text{LN TIME} = f(\text{LN FLOW}, \text{LN WT}, \text{LN } CO_2, \text{LN PRESS}, \text{LN RH}, \text{LN TEMP})$$

A standard error comparison shows that this multiplicative model has standard error of 0.38105 versus 0.39085 for the exponential model. R^2 = 0.924 and the residual plot of Figure 14 is approximately normal, while the plot of residuals versus predicted values is without structure, Figure 15.

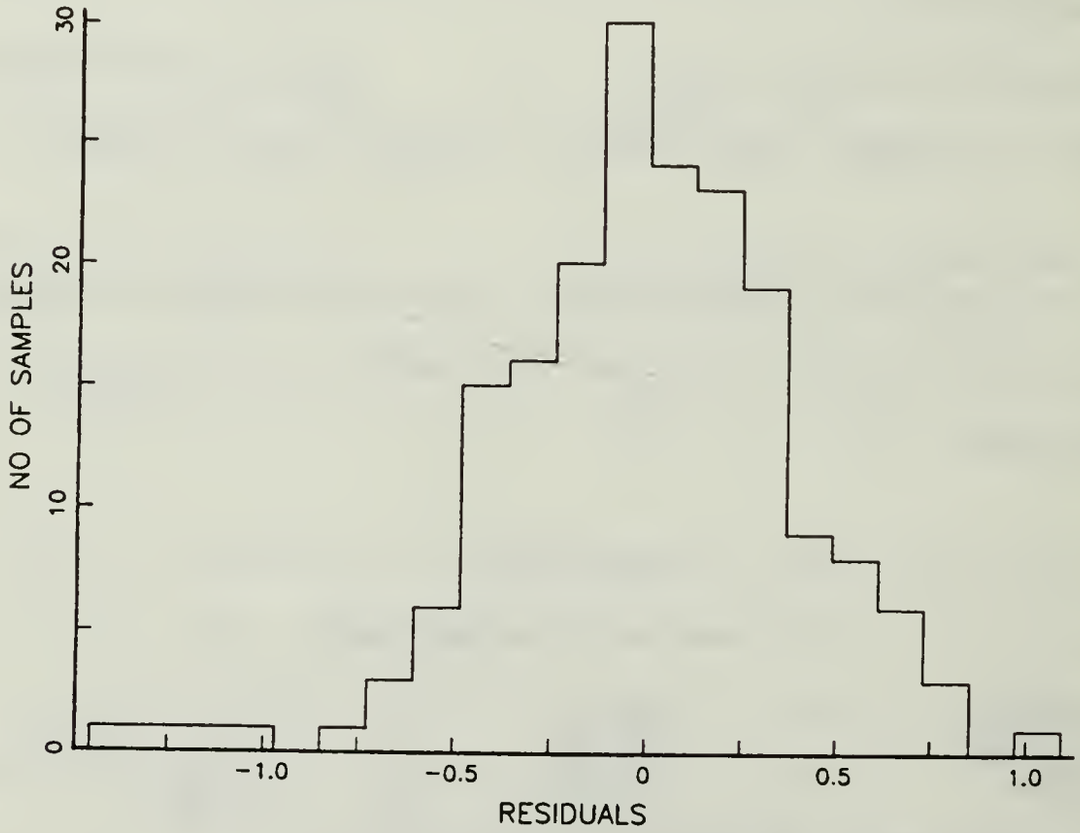


Figure 14. Histogram of Residuals

As with the exponential model, the plot of predicted values versus actual values generally supports the contention of common variance of the errors, Figure 16.

A final approach to modelling was made in an attempt to provide the capability to make predictions beyond the range

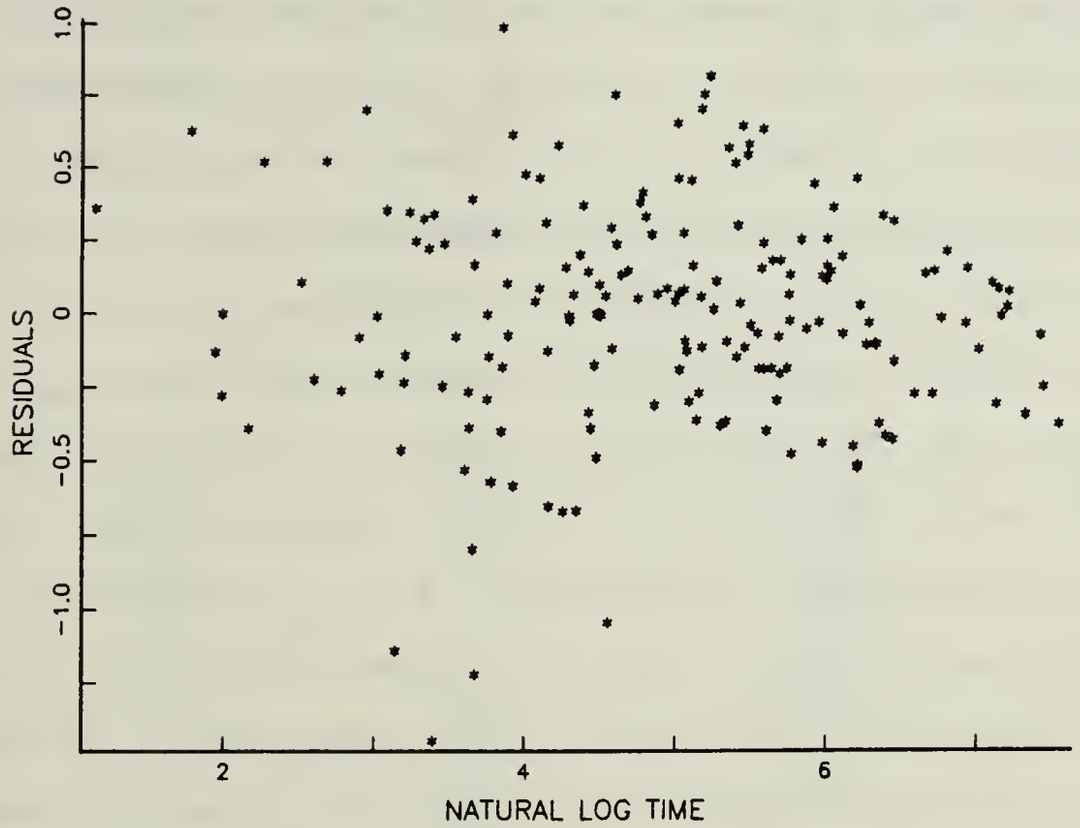


Figure 15. Plot of Residuals vs. Predicted

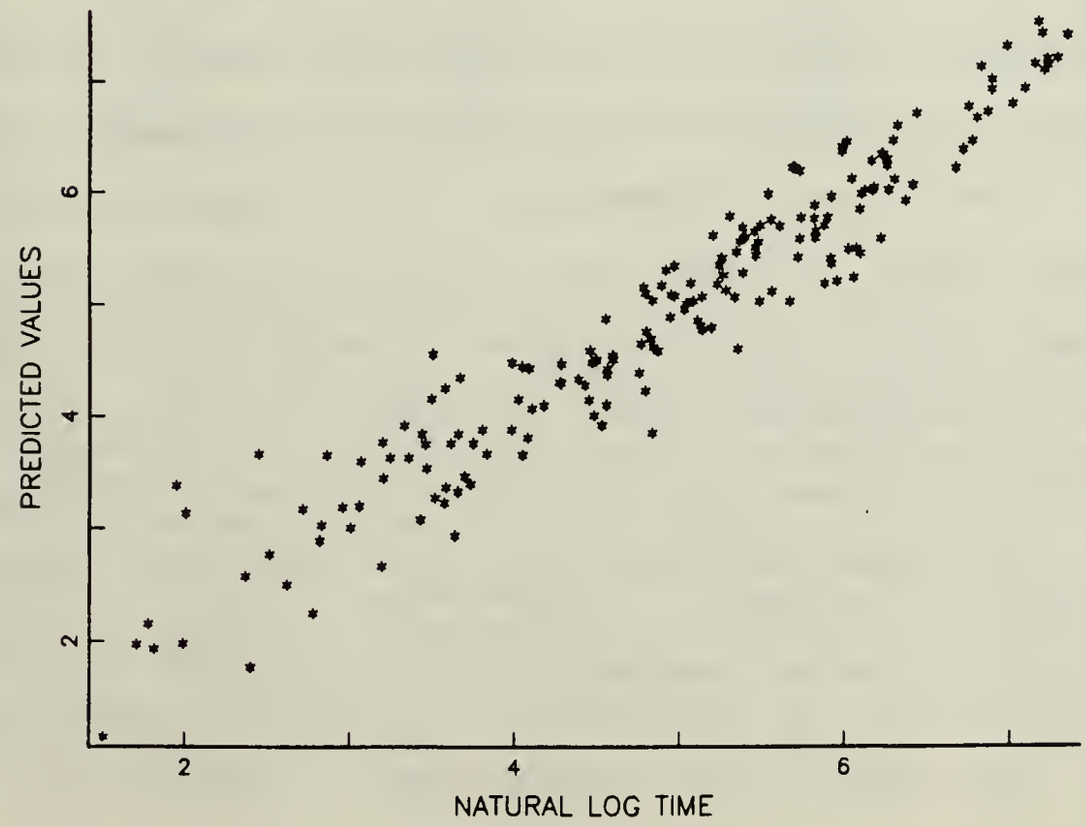


Figure 16. Plot of LN TIME vs. Predicted

of the current variables by using transformations of the predictor variables to redefine the data space. The result of the dimensional analysis discussed earlier was to identify efficiency as a function of five dimensionless groups. This expression can be rewritten to demonstrate that efficiency is a function of eight variables: $\eta = f(\rho,\bar{V},e,\mu,T,H,L,D)$. As defined earlier, the theoretical breakthrough time, t_{TH}, of a canister can be calculated by knowing the weight of absorbent, the flow rate, CO_2 concentration, and density of CO_2. Now, canister efficiency is defined as $\eta = t_B/t_{TH}$ and can be calculated for each trial or case in the data matrix.

This approach is to build a model from half of the original data set to predict efficiency, yielding t_B from the product of η and t_{TH}. From the data set, it is possible to express efficiency as a function of nine predictor variables: $\eta = f(RH,TEMP,WT,LENTH,DIAM,CO_2,FLOW,PRESS,HYDRA)$. Since efficiency is not affected by weight of the absorbent, WT may be removed from consideration. Of the remaining variables, only FLOW, LENTH, and DIAM are restrictive in the sense that they limit application of any model to small test size canisters. However, by reexpressing LENTH and DIAM as a ratio and replacing FLOW with mean linear gas velocity defined as $\bar{V} = FLOW/A_{cs}$, where A_{cs} is cross-sectional area of the canister, the predictor set ranges over the full extent of practical applications. Efficiency can now be expressed as a function of seven variables as, $\eta = f(RH,TEMP,L/D,CO_2,\bar{V},PRESS,HYDRA)$. Examination of the

correlation matrix of Table 13 reveals that several variables are negatively correlated, yet none are very highly correlated with each other.

TABLE 13

CORRELATION MATRIX

	RH	TEMP	LOD	CO_2
RH	1.000	-0.091	-0.061	0.087
TEMP	-0.091	1.000	0.035	-0.032
LOD	-0.061	0.035	1.000	-0.162
CO_2	0.087	-0.032	-0.162	1.000
VBAR	0.018	0.118	0.030	-0.113
PRESS	0.217	0.144	0.030	-0.417
HYDRA	-0.038	-0.071	-0.012	0.021
EFF	0.440	0.191	0.100	-0.275

	VBAR	PRESS	HYDRA	EFF
RH	0.018	0.217	-0.038	0.440
TEMP	0.118	0.144	-0.071	0.191
LOD	0.030	0.030	-0.012	0.100
CO_2	-0.113	-0.417	0.021	-0.275
VBAR	1.000	-0.024	-0.143	-0.362
PRESS	-0.024	1.000	0.097	0.183
HYDRA	-0.143	0.097	1.000	0.005
EFF	-0.362	0.183	0.005	1.000

After calculating efficiency as t_B/t_{TH}, it was discovered that in 17 of the 188 cases, efficiency was greater than one. These values were all changed to one as an efficiency greater than one has no meaning and in this situation is most likely the result of error in measurement. A regression analysis was run with efficiency as the dependent variable and resulted in an equation of the form

$$Y = \hat{\beta}_0 + \sum_{i=1}^{5} \hat{\beta}_i X_i$$

where the predictor variables are, in order, RH, $\overline{V}$, CO_2, TEMP, and PRESS. The variable coefficients are 0.00346, -0.64302, -0.11180, 0.00983, -0.00740 with a constant term of -0.10873. R^2 = 0.569 and all variables are significant at the α = 0.01 level. Both L/D and HYDRA were not selected as they were not significant at the 0.01 level. The residual plot of Figure 17 shows some structure suggesting that a transformation of data might provide a better fit.

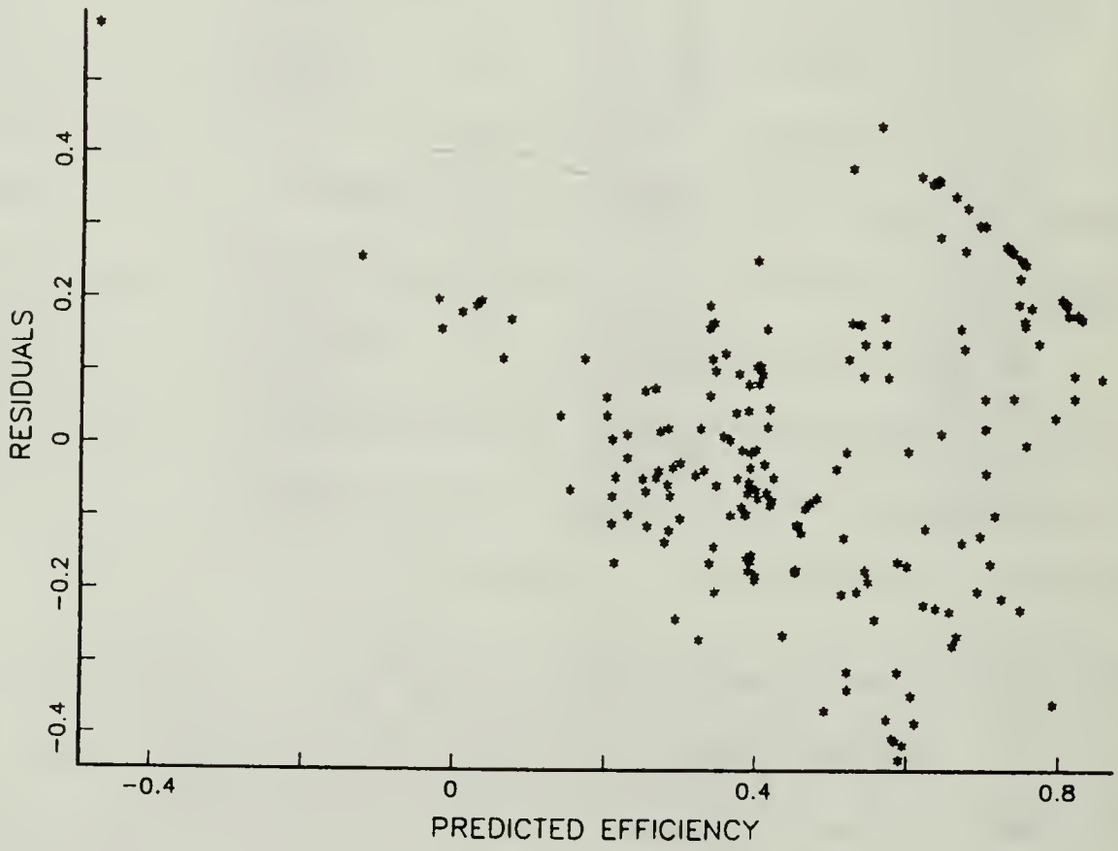


Figure 17. Plot of Residuals vs. Predicted

The histogram of residuals in Figure 18 indicates a slight departure from normality. Figure 19 is a plot of efficiency

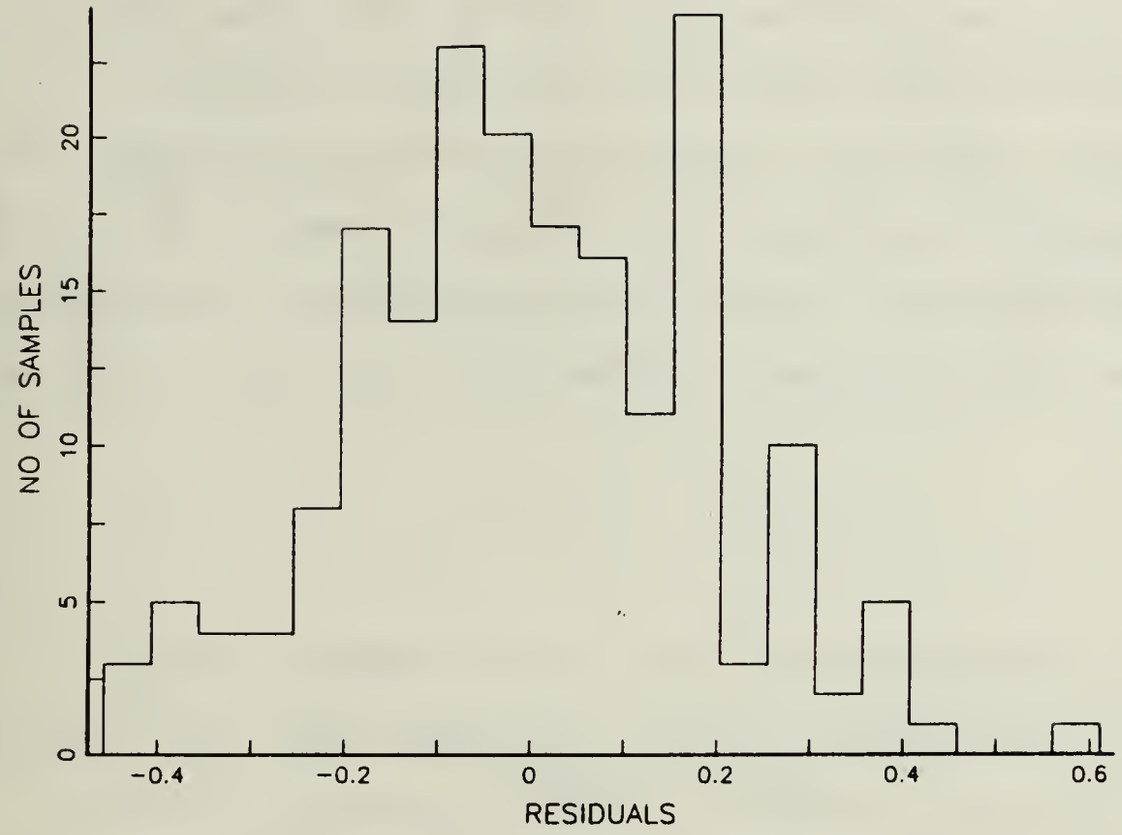


Figure 18. Histogram of Residuals

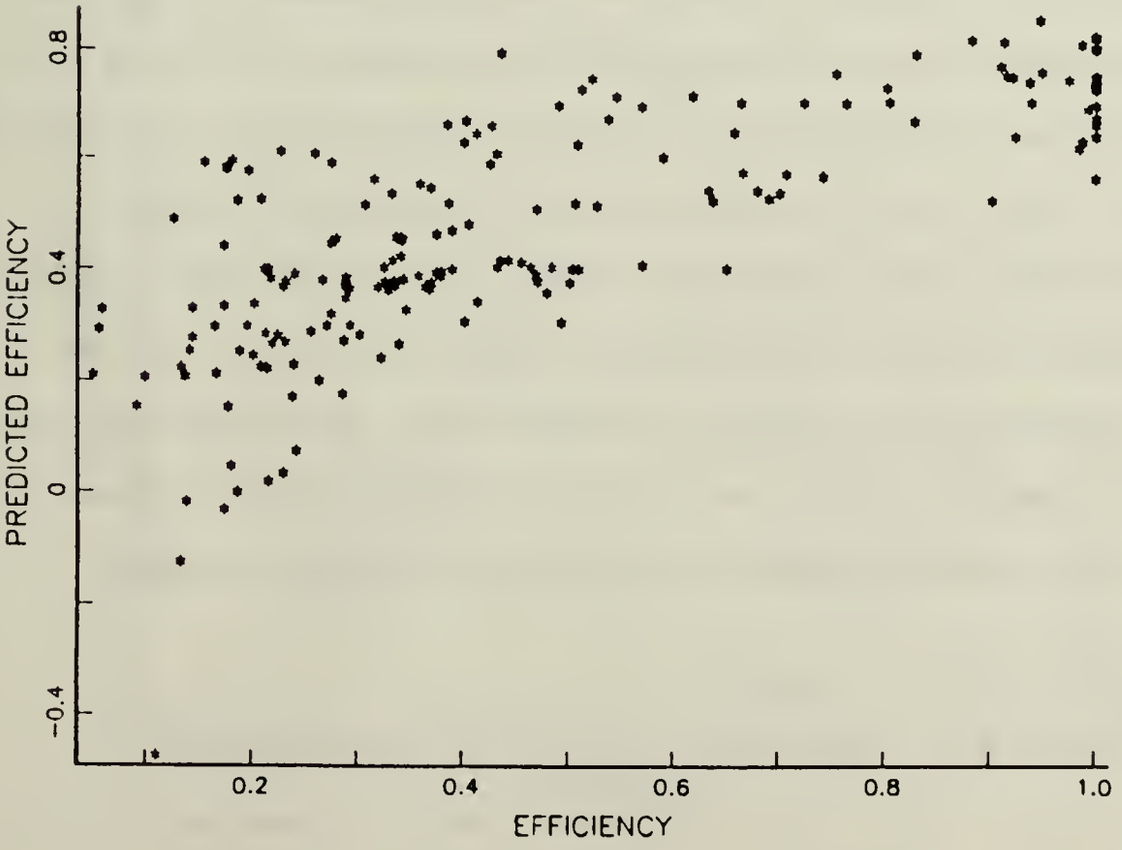


Figure 19. Plot of Efficiency vs. Predicted

versus predicted efficiency and depicts the variability of predictions over the range of the dependent variable.

The next step was to examine various transformations of the independent variables in an attempt to remove the structure from the residuals. Various trials using power transformations and cross terms resulted in a model for efficiency of the form

$$Y = \hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i X_i$$

where the independent variables are (RH × TEMP), (LN $\overline{V}$), (RH × CO_2), ($\overline{V}$ × PRESS), (RH × $\overline{V}$), and (TEMP × LOD). The variable coefficients are 0.0000816, -0.12947, -0.00141, -0.04283, -0.00297, and 0.000305 with a constant term of 0.01782. R^2 = 0.696 and all variables are significant at the α = 0.01 level. The residual plot of Figure 20 shows some improvement from the previous model and the histogram of residuals in Figure 21 is more normal than before. Figure 22 depicts the variability of predictions over the range of the dependent variable.

Both models were developed using half of the data set and cross-validation was performed as before using the second model, producing the results of Table 14. Shrinkage is equal to 0.156, however, R^2 is equal to 0.533 indicating a moderately strong relationship based on the scale provided earlier.

E. INTERPRETING THE RESULTS

Utilizing the techniques of multiple regression, two separate models have been developed, one exponential and the

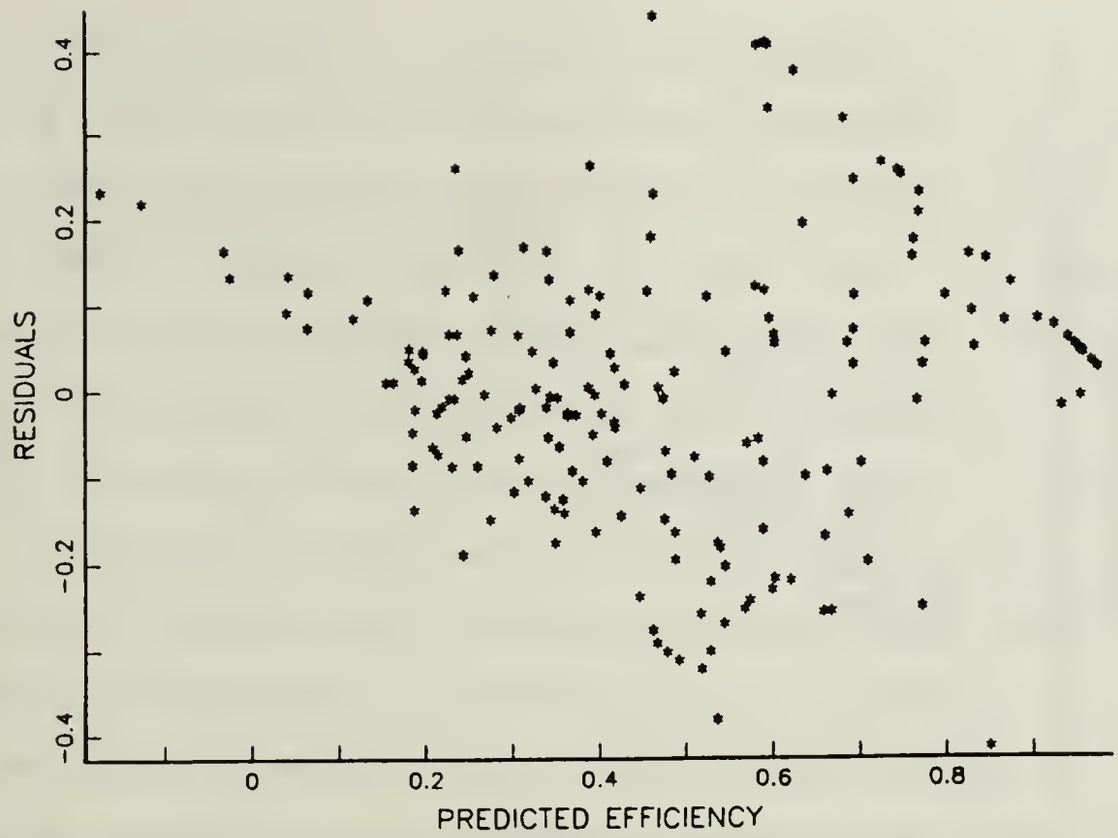


Figure 20. Plot of Residuals vs. Predicted

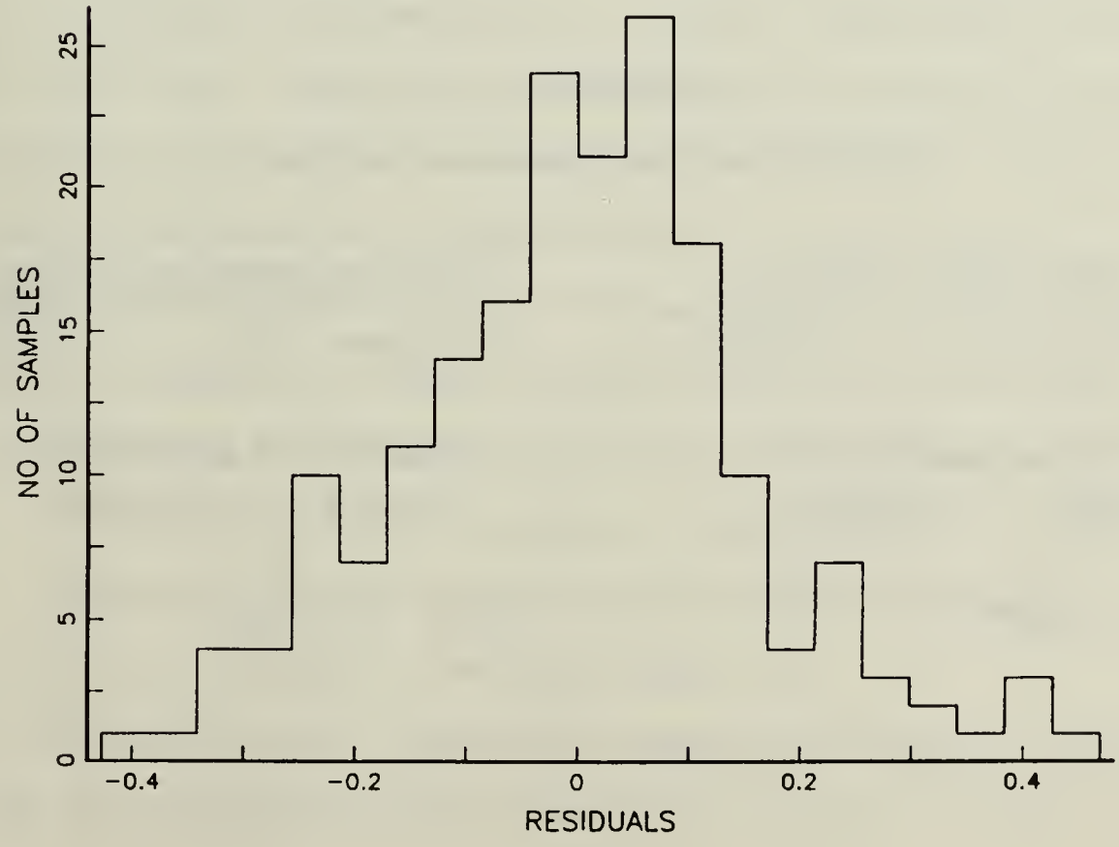


Figure 21. Histogram of Residuals

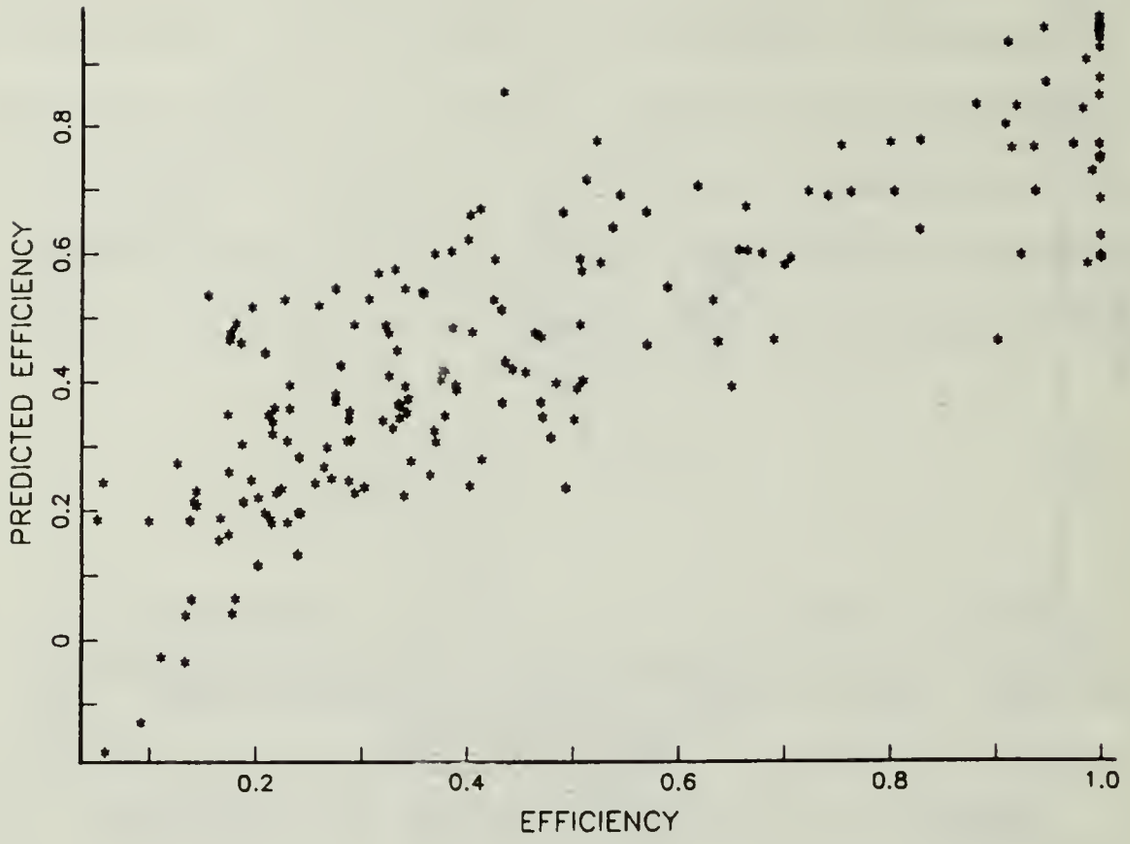


Figure 22. Plot of Efficiency vs. Predicted

TABLE 14

STATISTICS FOR CROSS-VALIDATION

Development Sample	Cross-Validation Sample
Variance of Y = 0.08655	Variance of Y = 0.07928
Variance of predicted values = 0.05989	Variance of predicted values = 0.04376
R^2 = 0.689	R^2_{CV} = 0.533

other multiplicative, to predict the canister life, t_B, of an axial flow canister containing High Performance Sodasorb. Both models have selected the same set of independent or predictor variables to estimate t_B. The variables chosen are percent relative humidity of the gas stream (RH), temperature in degrees Fahrenheit of the inlet gas (TEMP), weight in grams of the chemical absorbent (WT), percent by volume, surface level equivalent, of CO_2 in the inlet gas (CO_2), absolute rate of flow of the inlet gas in cubic centimeters per minute (FLOW), and absolute pressure of the environment expressed in atmospheres absolute (PRESS). The variables not included are canister length in inches (LENTH), canister diameter in centimeters (DIAM), and moisture level of the chemical absorbent (HYRDA). Earlier examination of the correlation matrix revealed that both length and diameter are highly correlated with both weight and flow. Chemical hydration level is usually fixed by the manufacturer and the results of this analysis indicate it has small significance in predicting t_B.

Both models have satisfied the assumptions inherent in regression analysis as demonstrated by the case analysis. All coefficients are highly significant and approximately 92% of the variability of the data is explained with each model. However, the worth or value of these models lies in their ability to predict canister life, t_B.

As stated earlier, regression analysis provides a basis for making predictions within the ranges of the values of the

independent or predictor variables. This restricts the use of the models to the ranges provided earlier in Table 4. If a canister of known dimensions is filled with HP Sodasorb and subjected to a gas stream flow rate with known CO_2 concentration and relative humidity at a particular depth and temperature, then the life of that canister can be predicted.

Predictions made outside the range of the data are suspect and not reliable. This is demonstrated by the following example calculations. The MK-12 surface supported dive system utilizes a canister 14.75 inches in length with a 6 inch inside diameter. For a dive depth of 390 feet (12.82 ATA), a flow rate of 6.0 absolute cuvic feet per minute (acfm) is required of which 0.6 is utilized by the diver leaving 5.4 acfm to enter the canister. A CO_2 level of 0.58% at the surface or 0.04524% at depth is assumed to enter the canister. The weight of absorbent inthe canister is 12 lbs. or 5448 grams, while the temperature is 60° F and the relative humidity is assumed to be 100%. Using the exponential model, t_B is predicted as 891 minutes or 14.85 hours. Theoretical canister breakthrough is calculated to be 1355 minutes or 22.6 hours. In actual practice, the MK-12 dive system has been utilized to slightly more than 10 hours as reported in Reference 1.

Since the models generated are restricted in use to the range of the values of the independent variables provided in the data set, this limits their application to small test size canisters that are too small for diving applications. In order

to successfully apply this analytical approach, additional tests would have to be made with larger canisters, allowing for greater absorbent weights and higher gas stream flow rates.

The final approach of modelling efficiency as a function of the data resulted in two models with R^2 values of 0.569 and 0.696 respectively. Most importantly, this approach allows t_B to be calculated from the product of η and t_{TH} over a much larger data space than before. Sample calculations were performed for the MK-12 dive profile as previously discussed for both models. The variable values were as follows: RH = 100, $\overline{V}$ = 0.46, CO_2 = 0.04524, TEMP = 60, PRESS = 12.82, L/D = 0.9678. The first model predicted an efficiency of 0.431 yielding a canister breakthrough time of 584 minutes or 9.7 hours. The second model predicted an efficiency of 0.230 yielding a canister breakthrough time of 311.7 minutes or 5.2 hours. Recall that canister breakthrough of approximately 10 hours has been observed in actual practice for this MK-12 dive profile. The ranges of the variables used in the model development are listed below, defining the space over which application of these models is restricted.

	MEAN	VAR	MIN	MAX
RH	61.7	2215	0	100
$\overline{V}$	0.2067	0.0446	0.0219	1.465
CO_2	1.1445	1.3295	0.031	4.35
TEMP	66.9947	84.6363	35	80
PRESS	4.1277	48.5291	1	32
L/D	2.6606	2.9050	0.2514	16.36

III. CONCLUSIONS

The CO_2 absorption process in a canister is a chemical process requiring both water and heat in order to perform. Water is provided in the chemical absorbent and in the moisture of the gas stream. The process is highly influenced by the rate of flow through the canister, the concentration of CO_2 in the inlet gas, the moisture content of the inlet gas, the operating temperature and, to some extent, the operating pressure. The effect of pressure is to change the density and viscosity of the inlet gas and more importantly to affect the actual water content of the gas stream [Ref. 1].

The models generated to predict canister breakthrough time appear to meet the assumptions inherent to multiple regression analysis and explain approximately 92% of the variability of the data. Their application is limited however to axial flow canisters with steady flow and, most importantly, to small canister dimensions. In order to utilize this approach of modelling canister breakthrough as a function of the available data and be able to apply the model to a wide range of canister sizes, more experimentation would be required.

Utilizing the definition of theoretical canister life and canister efficiency provided in Reference 1, canister efficiency was calculated for each of the 188 trials representing half of the data. Canister efficiency was then modelled as a

function of the eight independent variables. By reexpressing gas volumetric flow as gas mean velocity and reexpressing length and diameter as a ratio, the ranges of the independent variables in the development sample include parameter values for most dive profiles. The models for efficiency in Chapter II explain approximately 60% and 70% of the variability of the data, respectively. Simple cross-validation has shown that the models, while showing the expected shrinkage do exhibit good stability across different data sets.

These models should serve as an aid to the design process outlined in NCSC TECHMAN 4110. As models, they represent an attempt to quantify the relationship of canister performance with variuous environmental and geometric parameters.

APPENDIX A

DATA

RH	TEMP	WT	TIME	LENTH	DIAM	CO2	FLOW	PRESS	HYDRA
100	70	3.98	117.0	2.514	0.978	1.008	359.327	1.00	14.1
100	70	3.99	206.5	2.514	0.978	1.008	241.759	1.00	14.1
100	70	3.96	80.00	2.514	0.978	1.008	437.755	1.00	14.1
100	70	3.99	336.0	2.545	0.978	1.008	188.402	1.00	14.1
100	70	2.83	444.0	1.982	0.978	1.008	100.281	1.00	14.1
100	70	4.01	72.00	2.514	0.978	1.008	311.515	1.00	14.1
100	50	4.01	543.5	2.545	0.978	1.008	109.382	1.00	14.1
100	50	4.00	49.00	2.232	0.978	1.008	416.673	1.00	14.1
100	50	4.47	137.0	2.326	0.978	1.008	162.981	1.00	14.1
100	35	3.41	150.5	3.436	0.978	1.008	146.682	1.00	14.1
100	35	4.85	68.00	2.904	0.978	1.008	390.994	1.00	14.1
100	35	4.21	84.00	2.482	0.978	1.008	278.027	1.00	14.1
100	35	3.98	231.5	2.545	0.978	1.008	108.971	1.00	14.1
100	35	4.03	153.5	2.514	0.978	1.008	166.340	1.00	14.1
100	35	4.01	38.00	2.545	0.978	1.008	484.333	1.00	14.1
100	35	4.06	394.5	2.482	0.978	1.008	68.141	1.00	14.1
100	35	3.99	98.00	2.451	0.978	1.008	227.107	1.00	14.1
100	35	3.98	57.00	2.639	0.978	1.008	389.565	1.00	14.1
100	35	4.03	48.50	2.514	0.978	1.008	427.547	1.00	14.1
100	35	4.08	113.5	2.639	0.978	1.008	204.429	1.00	14.1
100	35	3.98	259.5	2.576	0.978	1.008	93.9988	1.00	14.1
100	35	4.05	68.50	2.670	0.978	1.008	316.940	1.00	14.1
100	35	3.98	50.00	2.420	0.978	1.008	390.902	1.00	14.1
100	35	3.98	40.00	2.676	0.978	1.008	471.154	1.00	14.1
100	50	4.01	60.00	2.514	0.978	1.008	327.174	1.00	14.1
100	50	4.02	112.0	2.514	0.978	1.008	196.598	1.00	14.1
100	50	4.03	262.0	2.514	0.978	1.008	95.6553	1.00	14.1
100	50	4.09	48.00	2.545	0.978	1.008	405.315	1.00	14.1
100	70	4.00	38.00	2.732	0.978	1.008	483.235	1.00	14.1
100	70	4.07	61.00	2.545	0.978	1.008	321.092	1.00	14.1
100	70	4.03	46.00	2.732	0.978	1.008	399.155	1.00	14.1
100	70	3.99	95.50	2.545	0.978	1.008	219.081	1.00	14.1
100	70	3.96	202.5	2.420	0.978	1.008	129.562	1.00	14.1
100	70	3.99	224.0	2.607	0.978	1.008	127.555	1.00	14.1
100	70	4.00	36.00	2.482	0.978	1.008	481.520	1.00	14.1
100	70	3.88	59.50	2.607	0.978	1.008	350.382	1.00	14.1
100	70	4.02	142.5	2.670	0.978	1.008	226.582	1.00	14.1
100	70	4.12	49.00	2.732	0.978	1.008	431.004	1.00	14.1
100	70	4.00	552.0	2.576	0.978	1.008	100.087	1.00	14.1
100	70	4.00	157.0	2.639	0.978	1.008	248.877	1.00	14.1
100	70	4.05	49.00	2.514	0.978	1.008	423.562	1.00	14.1
100	50	4.06	61.00	2.576	0.978	1.008	346.828	1.00	14.1
100	50	4.11	46.00	2.545	0.978	1.008	430.357	1.00	14.1
100	50	4.01	94.50	2.545	0.978	1.008	233.644	1.00	14.1
100	50	4.01	253.0	2.576	0.978	1.008	106.892	1.00	14.1
100	50	2.05	120.0	1.389	0.978	1.009	99.165	1.00	14.1
100	50	3.10	296.0	2.139	0.978	1.009	74.260	1.00	14.1
50	50	3.98	117.0	2.514	0.978	1.008	359.327	1.00	14.1
50	50	3.99	206.5	2.514	0.978	1.008	241.759	1.00	14.1
50	50	3.96	80.00	2.514	0.978	1.008	437.755	1.00	14.1
50	50	3.99	336.0	2.545	0.978	1.008	188.402	1.00	14.1
50	50	2.83	444.0	1.982	0.978	1.008	100.281	1.00	14.1
50	50	4.01	72.00	2.514	0.978	1.008	311.515	1.00	14.1
50	50	4.01	543.5	2.545	0.978	1.008	109.382	1.00	14.1
50	50	4.00	49.00	2.232	0.978	1.008	416.673	1.00	14.1
50	50	4.47	137.0	2.326	0.978	1.008	162.981	1.00	14.1
50	35	3.41	150.5	3.436	0.978	1.008	146.682	1.00	14.1
50	35	4.85	68.00	2.904	0.978	1.008	390.994	1.00	14.1
50	35	4.21	84.00	2.482	0.978	1.008	278.027	1.00	14.1
50	35	3.98	231.5	2.545	0.978	1.008	108.971	1.00	14.1
50	35	4.03	153.5	2.514	0.978	1.008	166.340	1.00	14.1
50	35	4.01	38.00	2.545	0.978	1.008	484.333	1.00	14.1
50	35	4.06	394.5	2.482	0.978	1.008	68.141	1.00	14.1
50	35	3.99	98.00	2.451	0.978	1.008	227.107	1.00	14.1
50	35	3.98	57.00	2.639	0.978	1.008	389.565	1.00	14.1
50	35	4.03	48.50	2.514	0.978	1.008	427.547	1.00	14.1
50	35	4.08	113.5	2.639	0.978	1.008	204.429	1.00	14.1
50	35	3.98	259.5	2.576	0.978	1.008	93.9988	1.00	14.1
50	35	4.05	68.50	2.670	0.978	1.008	316.940	1.00	14.1
50	35	3.98	50.00	2.420	0.978	1.008	390.902	1.00	14.1
50	35	3.98	40.00	2.676	0.978	1.008	471.154	1.00	14.1
50	50	4.01	60.00	2.514	0.978	1.008	327.174	1.00	14.1
50	50	4.02	112.0	2.514	0.978	1.008	196.598	1.00	14.1
50	50	4.03	262.0	2.514	0.978	1.008	95.6553	1.00	14.1
50	50	4.09	48.00	2.545	0.978	1.008	405.315	1.00	14.1
50	70	4.00	38.00	2.732	0.978	1.008	483.235	1.00	14.1
50	70	4.07	61.00	2.545	0.978	1.008	321.092	1.00	14.1
50	70	4.03	46.00	2.732	0.978	1.008	399.155	1.00	14.1
50	70	3.99	95.50	2.545	0.978	1.008	219.081	1.00	14.1
50	70	3.96	202.5	2.420	0.978	1.008	129.562	1.00	14.1
50	70	3.99	224.0	2.607	0.978	1.008	127.555	1.00	14.1
50	70	4.00	36.00	2.482	0.978	1.008	481.520	1.00	14.1
50	70	3.88	59.50	2.607	0.978	1.008	350.382	1.00	14.1
50	70	4.02	142.5	2.670	0.978	1.008	226.582	1.00	14.1
50	70	4.12	49.00	2.732	0.978	1.008	431.004	1.00	14.1
50	70	4.00	552.0	2.576	0.978	1.008	100.087	1.00	14.1
50	70	4.00	157.0	2.639	0.978	1.008	248.877	1.00	14.1
50	70	4.05	49.00	2.514	0.978	1.008	423.562	1.00	14.1
50	50	4.06	61.00	2.576	0.978	1.008	346.828	1.00	14.1
50	50	4.11	46.00	2.545	0.978	1.008	430.357	1.00	14.1
50	50	4.01	94.50	2.545	0.978	1.008	233.644	1.00	14.1
50	50	4.01	253.0	2.576	0.978	1.008	106.892	1.00	14.1
50	50	2.05	120.0	1.389	0.978	1.009	99.165	1.00	14.1
50	50	3.10	296.0	2.139	0.978	1.009	74.260	1.00	14.1

100	70	4.80	120.5	2.701	0.978	1.070	576.577	1.0	14.1
100	70	4.30	58.8	2.701	0.978	1.070	313.124	1.0	20.1
100	70	3.88	168.5	2.764	0.978	1.070	317.419	1.0	7.0
100	70	3.84	247.0	2.764	0.978	1.070	313.528	1.0	0.6
100	70	4.31	59.1	2.701	0.978	1.070	315.523	1.0	20.1
100	70	3.70	19.3	2.732	0.978	1.070	852.459	1.0	7.0
100	70	3.85	115.6	2.701	0.978	1.070	411.930	1.0	7.0
100	70	4.45	33.4	2.764	0.978	1.070	431.338	1.0	20.1
100	70	3.61	160.2	2.701	0.978	1.050	430.079	1.0	0.6
100	70	3.73	186.0	2.732	0.978	1.070	408.409	1.0	3.8
100	70	3.76	168.8	2.701	0.978	1.050	432.599	1.0	3.8
100	70	3.99	92.1	2.701	0.978	1.050	589.980	1.0	7.0
100	70	3.78	125.5	2.701	0.978	1.050	581.052	1.0	0.6
100	70	3.90	30.2	2.701	0.978	1.070	541.284	1.0	3.8
100	70	3.63	24.6	2.701	0.978	1.050	586.453	1.0	3.8
100	70	4.68	22.8	2.701	0.978	1.050	584.960	1.0	20.1
100	70	4.16	137.5	2.732	0.978	1.050	565.177	1.0	0.6
100	70	3.49	209.1	2.701	0.978	1.070	323.781	1.0	3.8
0	70	3.79	33.8	2.732	0.978	1.020	369.166	1.0	7.0
0	70	4.63	72.7	2.732	0.978	1.020	359.066	1.0	20.1
0	70	3.63	22.6	2.701	0.978	1.020	356.507	1.0	3.8
0	70	3.39	11.6	2.701	0.978	1.020	353.318	1.0	0.6
0	70	4.44	55.7	2.701	0.978	1.020	343.559	1.0	20.1
0	70	3.90	31.9	2.701	0.978	1.020	387.225	1.0	7.0
0	70	4.66	60.5	2.732	0.978	1.020	381.203	1.0	20.1
0	70	3.81	23.2	2.701	0.978	1.020	370.665	1.0	3.8
0	70	3.51	10.7	2.732	0.978	1.020	377.923	1.0	0.6
0	70	4.63	32.0	2.732	0.978	1.020	529.898	1.0	20.1
0	70	3.81	21.3	2.701	0.978	1.020	538.776	1.0	7.0
0	70	3.76	15.1	2.701	0.978	1.020	540.230	1.0	3.8
0	70	3.62	7.4	2.701	0.978	1.020	532.032	1.0	0.6
0	70	4.80	45.0	2.732	0.978	1.020	442.128	1.0	20.1
0	70	4.07	28.8	2.732	0.978	1.020	439.208	1.0	7.0
0	70	3.79	18.4	2.732	0.978	1.020	437.658	1.0	3.8
0	70	3.37	8.0	2.732	0.978	1.020	434.099	1.0	0.6
100	70	3.71	33.2	2.732	0.978	1.020	475.409	1.0	3.8
100	70	3.93	43.1	2.732	0.978	1.020	431.025	1.0	3.8
100	70	3.96	224.2	2.732	0.978	1.020	371.203	1.0	3.8
100	70	3.67	33.1	2.732	0.978	1.020	478.763	1.0	3.8
100	70	3.82	183.6	2.732	0.978	1.020	445.587	1.0	3.8
100	70	3.81	257.6	2.670	0.975	0.980	280.929	1.0	7.0
100	70	3.71	605.9	2.576	0.975	0.980	148.677	1.0	7.0

APPENDIX B

DATA STATISTICS

STATISTICS C1

MEAN: 56.76392573
VARIANCE: 2301.202099
STD. DEV.: 47.97084635
COEFF. OF VARIATION: 0.8450938819
LOWER QUARTILE: 0
UPPER QUARTILE: 100
MEDIAN: 100
TRIMEAN: 75
MIDMEAN: 56.76392573
RANGE: 100
MIDRANGE: 50
MEAN ABSOLUTE DEVIATION: 43.23607427
INTERQUARTILE RANGE: 100
COEFF. OF SKEWNESS: -0.2722239663
COEFF. OF KURTOSIS: -1.858103566

STATISTICS C2

MEAN: 66.47214854
VARIANCE: 91.77648287
STD. DEV.: 9.580004325
COEFF. OF VARIATION: 0.1441205758
LOWER QUARTILE: 70
UPPER QUARTILE: 70
MEDIAN: 70
TRIMEAN: 70
MIDMEAN: 70
RANGE: 45
MIDRANGE: 57.5
MEAN ABSOLUTE DEVIATION: 3.899204244
INTERQUARTILE RANGE: 0
COEFF. OF SKEWNESS: -2.264855315
COEFF. OF KURTOSIS: 4.098664453

STATISTICS C3

MEAN: 19.99413793
VARIANCE: 4049.837996
STD. DEV.: 63.63833747
COEFF. OF VARIATION: 3.182849778
LOWER QUARTILE: 3.99
UPPER QUARTILE: 4.8
MEDIAN: 4
TRIMEAN: 4.1975
MIDMEAN: 4.092277228
RANGE: 381.51
MIDRANGE: 192.805
MEAN ABSOLUTE DEVIATION: 16.10692308
INTERQUARTILE RANGE: 0.81
COEFF. OF SKEWNESS: 4.605224212
COEFF. OF KURTOSIS: 20.84155735

STATISTICS C4

MEAN: 261.4058355
VARIANCE: 102332.1454
STD. DEV.: 319.8939597
COEFF. OF VARIATION: 1.223744524
LOWER QUARTILE: 53.5
UPPER QUARTILE: 333.3
MEDIAN: 150.8
TRIMEAN: 172.1
MIDMEAN: 160.2650794
RANGE: 1966.6
MIDRANGE: 986.4
MEAN ABSOLUTE DEVIATION: 199.1527851
INTERQUARTILE RANGE: 279.8
COEFF. OF SKEWNESS: 2.267689597
COEFF. OF KURTOSIS: 5.635678376

STATISTICS C5

MEAN: 2.679376658
VARIANCE: 1.233382645
STD. DEV.: 1.110577618
COEFF. OF VARIATION: 0.4144910401
LOWER QUARTILE: 2.514
UPPER QUARTILE: 2.731
MEDIAN: 2.639
TRIMEAN: 2.63075
MIDMEAN: 2.627466667
RANGE: 14.953
MIDRANGE: 8.5235
MEAN ABSOLUTE DEVIATION: 0.2911564987
INTERQUARTILE RANGE: 0.217
COEFF. OF SKEWNESS: 9.754890674
COEFF. OF KURTOSIS: 110.1184256

STATISTICS C6

MEAN: 1.489450928
VARIANCE: 2.809112562
STD. DEV.: 1.67604074
COEFF. OF VARIATION: 1.125274226
LOWER QUARTILE: 0.978
UPPER QUARTILE: 0.978
MEDIAN: 0.978
TRIMEAN: 0.978
MIDMEAN: 0.978
RANGE: 8.471
MIDRANGE: 5.2105
MEAN ABSOLUTE DEVIATION: 0.5114827586
INTERQUARTILE RANGE: 0
COEFF. OF SKEWNESS: 3.78741941
COEFF. OF KURTOSIS: 13.6946683

STATISTICS C7

MEAN: 1.149448276
VARIANCE: 1.319485716
STD. DEV.: 1.148688694
COEFF. OF VARIATION: 0.9993391771
LOWER QUARTILE: 0.49
UPPER QUARTILE: 1.08
MEDIAN: 1.05
TRIMEAN: 0.9175
MIDMEAN: 1.000019512
RANGE: 4.319
MIDRANGE: 2.1905
MEAN ABSOLUTE DEVIATION: 0.6293050398
INTERQUARTILE RANGE: 0.59
COEFF. OF SKEWNESS: 1.916488355
COEFF. OF KURTOSIS: 2.948357655

STATISTICS C8

MEAN: 1039.887416
VARIANCE: 10965495.67
STD. DEV.: 3311.418981
COEFF. OF VARIATION: 3.184401435
LOWER QUARTILE: 132.905
UPPER QUARTILE: 437.658
MEDIAN: 269.806
TRIMEAN: 277.54375
MIDMEAN: 278.9265132
RANGE: 31161.903
MIDRANGE: 15604.7885
MEAN ABSOLUTE DEVIATION: 892.8289072
INTERQUARTILE RANGE: 304.753
COEFF. OF SKEWNESS: 5.535591624
COEFF. OF KURTOSIS: 34.31907269

STATISTICS C9

MEAN: 4.201591512
VARIANCE: 54.6241464
STD. DEV.: 7.390815002
COEFF. OF VARIATION: 1.759051298
LOWER QUARTILE: 1
UPPER QUARTILE: 2
MEDIAN: 1
TRIMEAN: 1.25
MIDMEAN: 1.074204947
RANGE: 31
MIDRANGE: 16.5
MEAN ABSOLUTE DEVIATION: 3.201591512
INTERQUARTILE RANGE: 1
COEFF. OF SKEWNESS: 2.768618901
COEFF. OF KURTOSIS: 7.042060544

STATISTICS C10

MEAN: 13.38594164
VARIANCE: 9.518977369
STD. DEV.: 3.085284001
COEFF. OF VARIATION: 0.2304868857
LOWER QUARTILE: 14.1
UPPER QUARTILE: 14.1
MEDIAN: 14.1
TRIMEAN: 14.1
MIDMEAN: 14.1
RANGE: 19.5
MIDRANGE: 10.35
MEAN ABSOLUTE DEVIATION: 1.000530504
INTERQUARTILE RANGE: 0
COEFF. OF SKEWNESS: -2.609146814
COEFF. OF KURTOSIS: 7.772805196

LIST OF REFERENCES

1. Naval Coastal Systems Center Techman 4110-1-83, Design Guidelines for Carbon Dioxide Scrubbers, by M.L. Nuckols, A. Purer, and G.A. Deason, May 1983.

2. Navy Department, U.S. Navy Diving Manual, NAVSEA 0994-LP-001-9010, Washington, D.C., 1978.

3. Navy Experimental Diving Unit Report 3-64, Carbon Dioxide Absorption Systems for SCUBA, by M.W. Goodman, 15 January 1965.

4. Navy Experimental Diving Unit Report 1-60, Carbon Dioxide Absorbent Evaluation and Canister Design, by H.W. Huseby and E.J. Michielsen, 6 November 1959.

5. Navy Experimental Diving Unit Report 9-57, Canister Design Criteria of Carbon Dioxide Removal from SCUBA, by G.J. Duffner, 8 March 1957.

6. The SODASORB Manual of Carbon Dioxide Absorption, W.R. Grace and Co., 1980.

7. A. Purer, G.A. Deason, and M.L. Nuckols, Carbon Dioxide Absorption Characteristics of Hydrated Calcium Hydroxide with Metal Hydroxide Activators, American Society of Mechanical Engineers, OED-Vol. 10, pp. 57-74, Winter 1982.

8. Shames, I.H., Mechanics of Fluids, pp. 188-204, McGraw-Hill, 1962.

9. Naval Coastal Systems Center TM 368-82, The Effects of Chemical Hydration Level on Carbon Dioxide Absorption by High Performance Sodasorb, by M.L. Nuckols, A. Purer, and G.A. Deason, November 1982.

10. Naval Coastal Systems Center TM 364-82, The Effects of Pulsatile Flow on Carbon Dioxide Absorption by High Performance Sodasorb, by A. Purer, G. Deason, R. Taylor, M. Nuckols, December 1982.

11. Weisberg, S., Applied Linear Regression, Wiley, 1980.

12. Younger, M.S., A Handbook for Linear Regression, Wadsworth, Inc., 1979.

13. Norusis, M.J., SPSS Introductory Guide: Basic Statistics and Operations, McGraw Hill, 1982.

14. Naval Aerospace Medical Research Laboratory, Monograph 17, The Influence of Selected Factors on Shirnkage and Overfit in Multiple Correlation, by N.E. Lane, September 1971.

15. Amick, D.J. and Walberg, H.J., Introductory Multivariate Analysis, McCutchan, 1975.

16. Mosteller, F. and Tukey, J., Data Analysis and Regression, Addison-Wesley, 1977.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943	2
3.	Professor C.W. Hutchins, Code 55Hw Department of Operations Research Naval Postgraduate School Monterey, California 93943	1
4.	Professor G.F. Lindsay, Code 55Ls Department of Operations Research Naval Postgraduate School Monterey, California 93943	1
5.	Dr. M.L. Nuckols, Code 4110 Naval Coastal Systems Center Panama City, Florida 32407	1
6.	LCDR J.E. Yarborough 2956 Ariane Dr. San Diego, California 92117	2

199-619

213243

Thesis
Y227
c.1

Yarborough
Models to predict the performance of axial flow carbon dioxide absorptive canisters.

3 JUN 86

31368

213243

Thesis
Y227
c.1

Yarborough
Models to predict the performance of axial flow carbon dioxide absorptive canisters.

thesY227
Models to predict the performance of axi
3 2768 000 61813 6
DUDLEY KNOX LIBRARY